UNIVERSITAT POLITÈCNICA DE VALÈNCIA

DEPARTAMENT DE MATEMÀTICA APLICADA

MEAN SQUARE SOLUTIONS OF RANDOM LINEAR MODELS AND COMPUTATION OF THEIR PROBABILITY DENSITY FUNCTION

AUTHOR

MARC JORNET SANZ

ADVISERS

PhD JUAN CARLOS CORTÉS LÓPEZ

PhD RAFAEL JACINTO VILLANUEVA MICÓ

València, December 2019

This thesis has been developed and written at the Institute of Multidisciplinary Mathematics (IMM), Universitat Politècnica de València, with the research group on *Dynamical systems with uncertainty and applications*. I was awarded a Research Staff Training (FPI) Grant 2017.

I obtained a grant for mobility from Universitat Politècnica de València for a research stay of three months (May–July 2019) at Centre de Mathématiques Appliquées (CMAP UMR7641), École Polytechnique, Palaiseau, France.

The work has been supported by the Spanish Ministerio de Economía y Competitividad grant MTM2017–89664–P, with project *Problemas Dinámicos con Incertidumbre Simulable: Modelización Matemática, Análisis, Computación y Aplicaciones*.

Abstract

This thesis concerns the analysis of differential equations with uncertain input parameters, in the form of random variables or stochastic processes with any type of probability distributions. In modeling, the input coefficients are set from experimental data, which often involve uncertainties from measurement errors. Moreover, the behavior of the physical phenomenon under study does not follow strict deterministic laws. It is thus more realistic to consider mathematical models with randomness in their formulation. The solution, considered in the sample-path or the mean square sense, is a smooth stochastic process, whose uncertainty has to be quantified. Uncertainty quantification is usually performed by computing the main statistics (expectation and variance) and, if possible, the probability density function.

In this dissertation, we study random linear models, based on ordinary differential equations with and without delay and on partial differential equations. The linear structure of the models makes it possible to seek for certain probabilistic solutions and even approximate their probability density functions, which is a difficult goal in general.

A very important part of the dissertation is devoted to random second-order linear differential equations, where the coefficients of the equation are stochastic processes and the initial conditions are random variables. The study of this class of differential equations in the random setting is mainly motivated because of their important role in Mathematical Physics. We start by solving the randomized Legendre differential equation in the mean square sense, which allows the approximation of the expectation and the variance of the stochastic

solution. The methodology is extended to general random second-order linear differential equations with analytic (expressible as random power series) coefficients, by means of the so-called Fröbenius method. A comparative case study is performed with spectral methods based on polynomial chaos expansions. On the other hand, the Fröbenius method together with Monte Carlo simulation are used to approximate the probability density function of the solution. Several variance reduction methods based on quadrature rules and multilevel strategies are proposed to speed up the Monte Carlo procedure. The last part on random second-order linear differential equations is devoted to a random diffusion-reaction Poisson-type problem, where the probability density function is approximated using a finite difference numerical scheme.

The thesis also studies random ordinary differential equations with discrete constant delay. We study the linear autonomous case, when the coefficient of the non-delay component and the parameter of the delay term are both random variables while the initial condition is a stochastic process. It is proved that the deterministic solution constructed with the method of steps that involves the delayed exponential function is a probabilistic solution in the Lebesgue sense.

Finally, the last chapter is devoted to the linear advection partial differential equation, subject to stochastic velocity field and initial condition. We solve the equation in the mean square sense and provide new expressions for the probability density function of the solution, even in the non-Gaussian velocity case.

Keywords: random differential equation, linear model, uncertainty quantification, probability density function, Fröbenius method, polynomial chaos expansions, Monte Carlo methods, random delayed equation, random linear advection partial differential equation.

Mathematics Subject Classification 2010: 34A30, 34F05, 35R60, 60H10, 60H15, 60H35, 65C05, 65C30.

Resum

Aquesta tesi tracta l'anàlisi d'equacions diferencials amb paràmetres d'entrada aleatoris, en la forma de variables aleatòries o processos estocàstics amb qualsevol mena de distribució de probabilitat. En modelització, els coeficients d'entrada són fixats a partir de dades experimentals, les quals solen comportar incertesa pels errors de mesurament. A més a més, el comportament del fenomen físic sota estudi no segueix patrons estrictament deterministes. És per tant més realista treballar amb models matemàtics amb aleatorietat en la seua formulació. La solució, considerada en el sentit de camins aleatoris o en el sentit de mitjana quadràtica, és un procés estocàstic suau, la incertesa del qual s'ha de quantificar. La quantificació de la incertesa és sovint duta a terme calculant els principals estadístics (esperança i variància) i, si es pot, la funció de densitat de probabilitat.

En aquest treball, estudiem models aleatoris lineals, basats en equacions diferencials ordinàries amb retard i sense, i en equacions en derivades parcials. L'estructura lineal dels models ens fa possible cercar certes solucions probabilístiques i inclús aproximar la seua funció de densitat de probabilitat, el qual és un objectiu complicat en general.

Una part molt important de la dissertació es dedica a les equacions diferencials lineals de segon ordre aleatòries, on els coeficients de l'equació són processos estocàstics i les condicions inicials són variables aleatòries. L'estudi d'aquesta classe d'equacions diferencials en el context aleatori està motivat principalment pel seu important paper en Física Matemàtica. Comencem resolent l'equació diferencial de Legendre aleatoritzada en el sentit de mitjana quadràtica, el que

permet l'aproximació de l'esperança i la variància de la solució estocàstica. La metodologia s'estén al cas general d'equacions diferencials lineals de segon ordre aleatòries amb coeficients analítics (expressables com a sèries de potències), per mitjà del conegut mètode de Fröbenius. Es duu a terme un estudi comparatiu amb mètodes espectrals basats en expansions de caos polinomial. Per altra banda, el mètode de Fröbenius juntament amb la simulació de Monte Carlo són emprats per a aproximar la funció de densitat de probabilitat de la solució. Per a accelerar el procediment de Monte Carlo, es proposen diversos mètodes de reducció de la variància basats en regles de quadratura i estratègies multinivell. L'última part sobre equacions diferencials lineals de segon ordre aleatòries estudia un problema aleatori de tipus Poisson de difusió-reacció, en què la funció de densitat de probabilitat és aproximada mitjançant un esquema numèric de diferències finites.

En la tesi també es tracten equacions diferencials ordinàries aleatòries amb retard discret i constant. Estudiem el cas lineal i autònom, quan el coeficient del component no retardat i el paràmetre del terme retardat són ambdós variables aleatòries mentre que la condició inicial és un procés estocàstic. Es prova que la solució determinista construïda amb el mètode dels passos i que involucra la funció exponencial retardada és una solució probabilística en el sentit de Lebesgue.

Finalment, el darrer capítol el dediquem a l'equació en derivades parcials lineal d'advecció, subjecta a velocitat i condició inicial estocàstiques. Resolem l'equació en el sentit de mitjana quadràtica i donem noves expressions per a la funció de densitat de probabilitat de la solució, inclús en el cas de velocitat no Gaussiana.

Paraules clau: equació diferencial aleatòria, model lineal, quantificació de la incertesa, funció de densitat de probabilitat, mètode de Fröbenius, expansió de caos polinomial, mètodes Monte Carlo, equació aleatòria amb retard, equació en derivades parcials lineal d'advecció aleatòria.

Classificació temàtica de matemàtiques 2010: 34A30, 34F05, 35R60, 60H10, 60H15, 60H35, 65C05, 65C30.

Resumen

Esta tesis trata el análisis de ecuaciones diferenciales con parámetros de entrada aleatorios, en la forma de variables aleatorias o procesos estocásticos con cualquier tipo de distribución de probabilidad. En modelización, los coeficientes de entrada se fijan a partir de datos experimentales, los cuales suelen acarrear incertidumbre por los errores de medición. Además, el comportamiento del fenómeno físico bajo estudio no sigue patrones estrictamente deterministas. Es por tanto más realista trabajar con modelos matemáticos con aleatoriedad en su formulación. La solución, considerada en el sentido de caminos aleatorios o en el sentido de media cuadrática, es un proceso estocástico suave, cuya incertidumbre se tiene que cuantificar. La cuantificación de la incertidumbre es a menudo llevada a cabo calculando los principales estadísticos (esperanza y varianza) y, si es posible, la función de densidad de probabilidad.

En este trabajo, estudiamos modelos aleatorios lineales, basados en ecuaciones diferenciales ordinarias con y sin retardo, y en ecuaciones en derivadas parciales. La estructura lineal de los modelos nos permite buscar ciertas soluciones probabilísticas e incluso aproximar su función de densidad de probabilidad, lo cual es un objetivo complicado en general.

Una parte muy importante de la disertación se dedica a las ecuaciones diferenciales lineales de segundo orden aleatorias, donde los coeficientes de la ecuación son procesos estocásticos y las condiciones iniciales son variables aleatorias. El estudio de esta clase de ecuaciones diferenciales en el contexto aleatorio está motivado principalmente por su importante papel en la Física Matemática.

Empezamos resolviendo la ecuación diferencial de Legendre aleatorizada en el sentido de media cuadrática, lo que permite la aproximación de la esperanza y la varianza de la solución estocástica. La metodología se extiende al caso general de ecuaciones diferenciales lineales de segundo orden aleatorias con coeficientes analíticos (expresables como series de potencias), mediante el conocido método de Fröbenius. Se lleva a cabo un estudio comparativo con métodos espectrales basados en expansiones de caos polinomial. Por otro lado, el método de Fröbenius junto con la simulación de Monte Carlo se utilizan para aproximar la función de densidad de probabilidad de la solución. Para acelerar el procedimiento de Monte Carlo, se proponen varios métodos de reducción de la varianza basados en reglas de cuadratura y estrategias multinivel. La última parte sobre ecuaciones diferenciales lineales de segundo orden aleatorias estudia un problema aleatorio de tipo Poisson de difusión-reacción, en el que la función de densidad de probabilidad es aproximada mediante un esquema numérico de diferencias finitas.

En la tesis también se tratan ecuaciones diferenciales ordinarias aleatorias con retardo discreto y constante. Estudiamos el caso lineal y autónomo, cuando el coeficiente de la componente no retardada y el parámetro del término retardado son ambos variables aleatorias mientras que la condición inicial es un proceso estocástico. Se demuestra que la solución determinista construida con el método de los pasos y que involucra la función exponencial retardada es una solución probabilística en el sentido de Lebesgue.

Finalmente, el último capítulo lo dedicamos a la ecuación en derivadas parciales lineal de advección, sujeta a velocidad y condición inicial estocásticas. Resolvemos la ecuación en el sentido de media cuadrática y damos nuevas expresiones para la función de densidad de probabilidad de la solución, incluso en el caso de velocidad no Gaussiana.

Palabras clave: ecuación diferencial aleatoria, modelo lineal, cuantificación de la incertidumbre, función de densidad de probabilidad, método de Fröbenius, expansión de caos polinomial, métodos Monte Carlo, ecuación aleatoria con retardo, ecuación en derivadas parciales lineal de advección aleatoria.

Clasificación temática de matemáticas 2010: 34A30, 34F05, 35R60, 60H10, 60H15, 60H35, 65C05, 65C30.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The important role played by differential equations in dealing with mathematical modeling is beyond discussion. They are powerful tools to describe the dynamics of phenomena appearing in a variety of distinct realms, such as Engineering, Biomedicine, Epidemiology, Chemistry, Social Sciences, etc. [13, 78, 86, 114, 145].

The differential equations governing the physical phenomena have inputs (initial and/or boundary conditions, forcing term and/or coefficients) that in practice need to be set from experimental data. These data often involve uncertainties from measurement errors, lack of information, etc. Moreover, the behavior of the physical phenomena does not follow strict deterministic laws. It is thus more realistic to consider mathematical models with randomness in their formulation. The model inputs are considered as random variables or stochastic processes rather than constants or deterministic functions, respectively. This approach leads to the area of random and stochastic differential equations. The solutions of such differential equations are stochastic processes.

A stochastic differential equation [1, 68, 109, 129] has the general differential form

$$\begin{cases} \mathrm{d}x(t) = b(t, x(t))\,\mathrm{d}t + \sigma(t, x(t))\,\mathrm{d}B(t), t \geq 0, \\ x(0) = x_0, \end{cases}$$

where $b, \sigma : [0, \infty) \times \mathbb{R} \to \mathbb{R}$ are given functions and $B(t)$ is a Brownian motion. Randomness is included into the deterministic model $x'(t) = \frac{\mathrm{d}x(t)}{\mathrm{d}t} = b(t, x(t))$ via an irregular perturbation, based on the formal derivative of the Brownian motion, $\xi(t) = B'(t)$, usually referred to as white noise process. Formally, $\xi(t)$ is a Gaussian process, with $\mathbb{E}[\xi(t)] = 0$ and $\mathbb{E}[\xi(t)\xi(s)] = \delta_0(t - s)$ (here $\mathbb{E}$ is the expectation and δ_0 is the Dirac delta function). Stochastic differential equations are studied through Itô calculus and model trajectories $x(t)$ with irregular patterns.

On the contrary, random differential equations arise to model smooth random phenomena, with not such irregular perturbations. The uncertainty, instead of being introduced by means of a white noise process, arises from imposing randomness into the input coefficients and the initial conditions, with any probability distribution: Normal, Gamma, Binomial, etc. Essentially, random differential equations arise from a direct randomization of their deterministic counterpart. The general form of a random differential equation initial value problem is the following:

$$\begin{cases} x'(t, \omega) = f(t, x(t, \omega), \omega),\ t \in I,\ \omega \in \Omega, \\ x(t_0, \omega) = x_0(\omega),\ \omega \in \Omega. \end{cases}$$

Here, $I \subseteq \mathbb{R}$ is an interval containing t_0, and Ω is the sample space of an underlying complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathcal{F} \subseteq 2^{\Omega}$ is the σ-algebra of events and $\mathbb{P}$ is the probability measure. The outcomes (i.e. the elements of Ω) are generically denoted by ω. The term $x(t, \omega)$ represents a stochastic process from $I \times \Omega$ to $\mathbb{R}^q$, and $x'(t, \omega) = \dot{x}(t, \omega) = \frac{\mathrm{d}}{\mathrm{d}t}x(t, \omega)$ represents its derivative in some probabilistic sense. We emphasize the fact that I is independent of ω.

This dissertation will be concerned with random differential equations, not with stochastic differential equations driven by irregular processes [157, pp. 96–98]. Although these two terms are sometimes treated as synonymous, they are conceptually different and require completely distinct techniques for the analysis.

Random differential equations are usually studied using two distinct approaches: sample-path calculus and L^p-calculus [124, 136, 160, 161, 167]. Essentially, sample-path sense means working with $x(\cdot, \omega) : I \to \mathbb{R}^q$ as a real map, for each $\omega \in \Omega$ fixed. While L^p-sense means to consider our stochastic solution as a map $x : I \to \mathrm{L}^p_q(\Omega)$ (or rather as an equivalence class), and thus work with continuity, differentiability, Riemann integrability, etc. in the topology of the random Lebesgue space $\mathrm{L}^p_q(\Omega)$. It is a classical result that any L^p-solution is also a sample-path solution [124, p. 140].

We recall that the Lebesgue space $\mathrm{L}^p(\Omega)$ is formed by the set of random variables $Y : \Omega \to \mathbb{R}$ such that $\|Y\|_p = \mathbb{E}[|Y|^p]^{1/p} < \infty$, $1 \leq p < \infty$, and $\|Y\|_\infty = \inf\{C > 0 : |Y| \leq C \text{ almost surely}\} < \infty$ (essential supremum norm). The Lebesgue space $\mathrm{L}^p_q(\Omega)$ consists of the random vectors $Y = (Y_1, \ldots, Y_q) : \Omega \to \mathbb{R}^q$ such that $\|Y\|_p = \max_{1\leq i\leq q} \|Y_i\|_p < \infty$. This space is Banach. In the case of $p = 2$ it is Hilbert, and it corresponds to the random variables having finite variance. The L^2_q-calculus is usually referred to as mean square calculus. A key feature of mean square calculus is that mean square convergence ensures convergence of the expectation and the variance. The L^4_q-calculus is often regarded as mean fourth calculus. Given the stochastic process x, we say that x is L^p-continuous at $t_1 \in I$ if $\lim_{h\to 0} \|x(t_1 + h) - x(t_1)\|_p = 0$. We say that x is L^p-differentiable at $t_1 \in I$ if $\lim_{h\to 0} \|\frac{x(t_1+h)-x(t_1)}{h} - x'(t_1)\|_p = 0$, for certain random variable/vector $x'(t_1)$ (called the derivative of x at t_1). If $I = [a, b]$, we say that x is L^p-Riemann integrable on $[a, b]$ if there exists a sequence of partitions $\{P_n\}_{n=1}^\infty$ with mesh tending to 0, $P_n = \{a = t_0^n < t_1^n < \ldots < t_{r_n}^n = b\}$, such that, for any choice of points $s_i^n \in [t_{i-1}^n, t_i^n]$, $i = 1, \ldots, r_n$, the limit $\lim_{n\to\infty} \sum_{i=1}^{r_n} x(s_i^n)(t_i^n - t_{i-1}^n)$ exists in $\mathrm{L}^p_q(\Omega)$; in this case, these Riemann sums have the same limit, which is a random variable/vector and is denoted by $\int_a^b x(t)\,\mathrm{d}t$. Finally, the stochastic process x is L^p-analytic at t_1 if $x(t) = \sum_{n=0}^\infty x_n(t - t_1)^n$ for every t in a neighborhood of t_1, where $x_0, x_1, \ldots$ are random variables/vectors and the sum is in the topology of $\mathrm{L}^p_q(\Omega)$.

Several methods have been proposed in the literature to obtain statistical information of the stochastic solution (uncertainty quantification).

Monte Carlo simulation is a popular statistical method based on obtaining independent realizations of the solution, $x(t, \omega_i)$, by solving the deterministic version of the problem for different realizations of the coefficients. This allows approximating the statistical moments of $x(t)$ from its generated sample [70]. Although it is a robust and easy to implement approach, this technique is computationally expensive (the root mean square error in the approximation of the expectation $\mathbb{E}[x(t)]$ is $\mathcal{O}(1/\sqrt{M})$, where M is the finite number of realizations).

Another method consists in using spectral expansions of the solution in terms of the random inputs, which present rapid mean square convergence. Polynomial chaos (PC) methods expand the solution in terms of Hermite polynomials when the inputs are Gaussian [169]. This method was extended using generalized polynomial chaos (gPC) expansions, to deal with non-Gaussian inputs [34, 51, 172, 173]. When the inputs of the differential equation are not independent, polynomial expansions based on the canonical polynomial basis were

proposed [49]. Essentially, the solution is approximated in the mean square sense as $x(t,\omega) \approx \sum_{i=1}^{N} \hat{x}_i(t)\phi_i(\zeta(\omega))$, where the $\hat{x}_i(t)$ are deterministic functions, $\zeta(\omega)$ is the vector of random input coefficients, and the ϕ_i are suitably selected multivariate polynomials.

For specific random differential equation problems, one may seek a mean square convergent random power series solution. This is the extension of the deterministic Fröbenius method to the random setting, which shows important applications for solving second-order linear differential equations [33]. Several differential equations from Mathematical Physics have been randomized and rigorously solved using the random Fröbenius method [27, 29, 43].

A more ambitious objective is the computation of the probability density function of $x(t)$, denoted as $f_{x(t)}(x) = \frac{\mathrm{d}(\mathbb{P}\circ x(t)^{-1})(x)}{\mathrm{d}x}$. The probability density function is defined as a non-negative Borel measurable function characterized by $\mathbb{P}[x(t) \in \mathcal{C}] = \int_{\mathcal{C}} f_{x(t)}(x)\,\mathrm{d}x$ (where $\mathcal{C} \subseteq \mathbb{R}^q$ is any Borel set). Random variables having a probability density function are called absolutely continuous, meaning that their probability law is absolutely continuous with respect to the Lebesgue measure. The density function allows calculating general statistics (expectation, variance, skewness, kurtosis, median, quantiles, mode, etc., provided these statistical quantities exist) and confidence intervals via integration.

This thesis aims at studying several random linear models, based on ordinary, delayed and partial differential equations. The linear structure of the models allows seeking for certain probabilistic solutions in the Lebesgue sense and even approximate their probability density functions, which is a difficult goal in general.

The structure of this dissertation, which consists of eleven chapters, is described in what follows:

Chapters 2–9 concern second-order random linear differential equations, where the coefficients of the equation are stochastic processes and the two initial conditions are random variables. The study of this class of differential equations in the random setting is mainly motivated on account of their important role in Mathematical Physics. Relevant equations in this regard are Airy, Hermite, Laguerre and Legendre random differential equations. We start in Chapter 2 solving the randomized Legendre differential equation in the mean square sense, which allows the approximation of the expectation and the variance of the stochastic solution. Chapter 3 generalizes the previous chapter to general second-order linear random differential equations with analytic coeffi-

cients (expressible as random power series). Chapter 4 relaxes the hypotheses from the previous chapter for particular equations, namely Airy, Hermite and Laguerre. The basis of these three chapters is the Fröbenius method, and rapid approximations for the mean and the variance of the output are obtained. In Chapter 5, the approximations of the statistics are constructed through gPC expansions. Chapter 6 aims at approximating the probability density function of the stochastic solution to general second-order linear random differential equations with analytic coefficients, by using Monte Carlo simulation methods. Chapter 7 proposes algorithmic improvements for the Monte Carlo methods of the previous chapter, based on path-wise selections, quadrature rules and multilevel Monte Carlo strategies. In Chapter 8, we study the random Legendre differential equation at a regular-singular point, by constructing a mean square solution based on random power series. Finally, Chapter 9 deals with a randomized diffusion-reaction Poisson-type problem, where a finite difference scheme is utilized to approximate the probability density function of the stochastic solution.

On the other hand, Chapter 10 is devoted to the random autonomous linear differential equation with discrete constant delay, in the case that the two coefficients (delay and non-delay term) are random variables and the initial condition is a stochastic process. The mean square solution is constructed using the method of steps and the delayed exponential function, and approximations of the mean and the variance are obtained.

In the last chapter, Chapter 11, we analyze the linear advection partial differential equation subject to stochastic velocity field and initial condition. We solve the equation in the mean square sense and estimate the probability density function of the stochastic solution.

Chapter 2

Improving the approximation of the first and second order statistics of the response process to the random Legendre differential equation

In this chapter, we deal with uncertainty quantification for the random Legendre differential equation, with input coefficient A and initial conditions X_0 and X_1. In a previous study [G. Calbo et al. In: Comput. Math. Appl. 61.9 (2011), pp. 2782–2792], a mean square convergent power series solution on $(-1/\mathrm{e}, 1/\mathrm{e})$ was constructed, under the assumptions of mean fourth integrability of X_0 and X_1, independence, and at most exponential growth of the absolute moments of A. In this chapter, we relax these conditions to construct an L^p solution ($1 \leq p \leq \infty$) to the random Legendre differential equation on the whole domain $(-1, 1)$, as in its deterministic counterpart. Our hypotheses assume no independence and less integrability of X_0 and X_1. Moreover, the growth condition on the moments of A is characterized by the boundedness of A, which simplifies the proofs significantly. We also provide approximations of the expectation and variance of the response process. The numerical experiments show the wide applicability of our findings. A comparison with Monte Carlo simulation is performed.

2.1 Introduction

The Fröbenius method has been successfully used to deal with particular second-order random linear differential equations. Without entering into the details in this first chapter, we mention Airy [43], Hermite [29], Legendre [31] and Bessel [46] differential equations. In this chapter, we will only deal with the random Legendre differential equation:

$$\begin{cases} (1-t^2)\ddot{X}(t) - 2t\dot{X}(t) + A(A+1)X(t) = 0, \ |t| < 1, \\ X(0) = X_0, \\ \dot{X}(0) = X_1. \end{cases} \tag{2.1}$$

The coefficient A is a non-negative random variable and the initial conditions X_0 and X_1 are random variables. All of them are defined in a common complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$. In the following chapters, we will extend the discussion to general random second-order linear differential equations with coefficients expressible as random power series (i.e. analytic).

Recall that, given a stochastic process $Y(t)$ with the property that $Y(t) \in \mathrm{L}^p(\Omega)$ for each t, one can define the $\mathrm{L}^p(\Omega)$ continuity, differentiability or analiticity of $Y(t)$, by taking limits in $\mathrm{L}^p(\Omega)$. This $\mathrm{L}^p(\Omega)$ random calculus is the setting in which one usually considers random differential equations such as (2.1). The particular case $p = 2$, which arises from working in the Hilbert space $\mathrm{L}^2(\Omega)$ and with random variables with well-defined expectation $\mathbb{E}[\cdot]$ and variance $\mathbb{V}[\cdot]$, is the most extended in the literature, and it is usually referred to as mean square calculus. An exposition of these topics is presented, for instance, in [160, 167].

In [31], the authors constructed a mean square convergent power series solution $X(t)$ to (2.1) on $(-1/\mathrm{e}, 1/\mathrm{e})$ under certain assumptions on the random inputs A, X_0 and X_1. The goal of this chapter is to improve [31]: to weaken the hypotheses from [31], to simplify the proofs significantly and to obtain an $\mathrm{L}^p(\Omega)$ random power series solution on the whole domain $(-1, 1)$, as in the deterministic counterpart of (2.1). Numerical examples that could not be tackled via the hypotheses from [31] will be carried out in this chapter, establishing a comparison with Monte Carlo simulation.

The structure of this chapter is the following. In Section 2.2, we will review the techniques used in [31]. We will relax the assumptions from [31] and we will improve the conclusions of the results. In Section 2.3, we will show how to approximate the expectation and variance of the response process, under no independence assumption. In Section 2.4, we will perform a wide

variety of examples and illustrate the potentiality of our findings by comparing the numerical results with Monte Carlo simulation. Section 2.5 will draw conclusions.

2.2 Random Legendre differential equation

In [31], the authors constructed a mean square power series solution to the random Legendre differential equation (2.1) on the time interval $(-1/\mathrm{e}, 1/\mathrm{e})$. The hypotheses assumed in [31] were that the absolute moments of A increased at most exponentially, that is, there exist two positive constants H and M such that

$$\mathbb{E}[|A|^n] \leq HM^n, \quad n \geq n_0; \tag{2.2}$$

that A is independent of the initial conditions X_0 and X_1; and that $X_0, X_1 \in \mathrm{L}^4(\Omega)$. Hypothesis (2.2) has been of constant use in the extant literature to study significant linear random differential equations via the Fröbenius method: [29, 31, 43]. In [31], the explicit solution to (2.1) was obtained in the form of a random power series solution by means of the Fröbenius method:

$$X(t) = X_0\tilde{X}_1(t) + X_1\tilde{X}_2(t) \tag{2.3}$$

for $|t| < 1/\mathrm{e}$, where

$$\tilde{X}_1(t) = \sum_{m=0}^{\infty} \frac{(-1)^m}{(2m)!} P_1(m) t^{2m}, \quad \tilde{X}_2(t) = \sum_{m=0}^{\infty} \frac{(-1)^m}{(2m+1)!} P_2(m) t^{2m+1}, \tag{2.4}$$

$$P_1(m) = \prod_{k=1}^{m} (A-2k+2)(A+2k-1), \quad P_2(m) = \prod_{k=1}^{m} (A-2k+1)(A+2k). \tag{2.5}$$

The series in (2.4) were proved to be mean fourth convergent for $|t| < 1/\mathrm{e}$. Since $X_0, X_1 \in \mathrm{L}^4(\Omega)$, it follows that (2.3) is a mean square solution to (2.1) on $(-1/\mathrm{e}, 1/\mathrm{e})$.

To summarize, the main result obtained in [31, Th. 11] was stated as follows:

Theorem 2.1 *Suppose that $X_0, X_1 \in \mathrm{L}^4(\Omega)$, that A satisfies the growth condition (2.2), and that A is independent of X_0 and X_1. Then the stochastic process defined by (2.3)–(2.5) is a mean square solution to the random initial value problem (2.1) on the time domain $(-1/\mathrm{e}, 1/\mathrm{e})$.*

Our goal is to extend this theorem and to simplify its proof given in [31]. The growth condition (2.2) was established in order to demonstrate the mean

fourth convergence of (2.4), by applying well-known inequalities: Hölder's inequality, c_s-inequality and arithmetic-geometric inequality. We will simplify the proof given in [31] by working with an equivalent but easier to manage form of (2.2), see Lemma 2.2. Moreover, the $\mathrm{L}^\infty(\Omega)$ convergence of (2.4) (which implies mean fourth convergence) will be obtained on the whole interval $(-1,1)$, see Theorem 2.4. This will provide the complete extension of the deterministic counterpart for the random Legendre differential equation.

Lemma 2.2 *The growth condition (2.2) is equivalent to the boundedness of A:* $\|A\|_\infty < \infty$.

Proof. If $\|A\|_\infty < \infty$, then $E[|A|^n] \leq \|A\|_\infty^n$, so that we can take $H = 1$ and $M = \|A\|_\infty$ and (2.2) is satisfied.
On the other hand, if (2.2) holds, then $\|A\|_n \leq H^{1/n} M$. By taking limits, $\|A\|_\infty = \lim_{n\to\infty} \|A\|_n \leq M < \infty$.

□

Lemma 2.3 *Let* $X(t) = \sum_{n=0}^\infty X_n t^n$ *be a formal random power series on* $(-1,1)$. *Let* $1 \leq p \leq \infty$. *Then the given series converges in* $\mathrm{L}^p(\Omega)$ *for all* $t \in (-1,1)$, *if and only if* $\sum_{n=0}^\infty \|X_n\|_p |t|^n < \infty$ *for all* $t \in (-1,1)$.

Proof. If $\sum_{n=0}^\infty \|X_n\|_p |t|^n < \infty$ for all $t \in (-1,1)$, then the series converges in $\mathrm{L}^p(\Omega)$ for all $t \in (-1,1)$, because in a Banach space, absolute convergence of a series implies convergence.
On the other hand, suppose that the series converges in $\mathrm{L}^p(\Omega)$ for all $t \in (-1,1)$. Fix $|t_0| < 1$. Let $|t_0| < |\rho| < 1$. Since $\sum_{n=0}^\infty \|X_n\|_p |\rho|^n < \infty$, then $\|X_n\|_p |\rho|^n \leq 1$, for $n \geq n_0$. Thus, $\|X_n\|_p |t_0|^n \leq (|t_0|/|\rho|)^n$, for $n \geq n_0$, with $\sum_{n=0}^\infty (|t_0|/|\rho|)^n < \infty$. By comparison, $\sum_{n=0}^\infty \|X_n\|_p |t_0|^n < \infty$.

□

We state and prove the main Theorem 2.4. It is a significant improvement of Theorem (2.1) stated and proved in [31]: for $p = 2$, we only require mean square integrability of X_0 and X_1, not mean fourth integrability; we do not need any independence assumption on A, X_0 and X_1; and we demonstrate mean square convergence of the series on the whole interval $(-1,1)$, not just $(-1/\mathrm{e}, 1/\mathrm{e})$. Moreover, our proof is much simpler, because the hypothesis of boundedness for A instead of the equivalent growth condition (2.2) allows simpler and more direct inequalities (we do not need Hölder's inequality, c_s-inequality, arithmetic-geometric inequality, etc.).

Theorem 2.4 *Suppose that $X_0, X_1 \in \mathrm{L}^p(\Omega)$, for certain $1 \leq p \leq \infty$, and $\|A\|_\infty < \infty$. Then the stochastic process defined by (2.3)–(2.5) is the unique $\mathrm{L}^p(\Omega)$ solution to the random initial value problem (2.1) on the whole time domain $(-1,1)$.*

Proof. From (2.3) and $X_0, X_1 \in \mathrm{L}^p(\Omega)$, it suffices to see that the two series given in (2.4) converge in $\mathrm{L}^\infty(\Omega)$ for $t \in (-1,1)$. That is,

$$\sum_{m=0}^{\infty} \frac{1}{(2m)!}\|P_1(m)\|_\infty |t|^{2m} < \infty, \ \sum_{m=0}^{\infty} \frac{1}{(2m+1)!}\|P_2(m)\|_\infty |t|^{2m+1} < \infty, \quad (2.6)$$

for $t \in (-1,1)$ (see Lemma 2.3). We will check (2.6) for the first series, as for the second one the reasoning is completely analogous.
Let $L = \|A\|_\infty$. We have

$$\begin{aligned}
\|P_1(m)\|_\infty &= \left\| \prod_{k=1}^{m} (A-2k+2)(A+2k-1) \right\|_\infty \leq \prod_{k=1}^{m} (L+2k-2)(L+2k-1) \\
&\leq \prod_{k=1}^{m} (L+2k-1)^2 = \left(\frac{\prod_{k=1}^{2m-1}(L+k)}{\prod_{k=1}^{m-1}(L+2k)} \right)^2 = \left(\frac{(L+2m-1)!}{L!\prod_{k=1}^{m-1}(L+2k)} \right)^2 \\
&= \left(\frac{(L+2m-1)!}{L!2^{m-1}\prod_{k=1}^{m-1}(L/2+k)} \right)^2 = \left(\frac{(L+2m-1)!\Gamma(L/2+1)}{L!2^{m-1}\Gamma(L/2+m)} \right)^2,
\end{aligned}$$

where the property $\Gamma(x) = (x-1)\Gamma(x-1)$ of the Gamma function $\Gamma(z) = \int_0^\infty x^{z-1}\mathrm{e}^{-x}\,\mathrm{d}x$ has been used. By the root test, if we check that

$$\lim_{m\to\infty} \left(\frac{(L+2m-1)!\Gamma(L/2+1)}{L!2^{m-1}\Gamma(L/2+m)(2m)!^{\frac{1}{2}}} \right)^{2/m} = 1,$$

then the first part of (2.6) will follow. By Stirling's formula, as $x \to \infty$, the asymptotic behavior of the Gamma function is $\Gamma(x) \sim \sqrt{2\pi x}(\frac{x-1}{\mathrm{e}})^{x-1}$. As a consequence,

$$\begin{aligned}
&\lim_{m\to\infty} \left(\frac{(L+2m-1)!\Gamma(L/2+1)}{L!2^{m-1}\Gamma(L/2+m)(2m)!^{\frac{1}{2}}} \right)^{2/m} \\
&= \lim_{m\to\infty} \left(\frac{\sqrt{2\pi(L+2m-1)}\left(\frac{L+2m-1}{\mathrm{e}}\right)^{L+2m-1}\Gamma(L/2+1)}{L!2^{m-1}\sqrt{2\pi(L/2+m)}\left(\frac{L/2+m-1}{\mathrm{e}}\right)^{L/2+m-1}\sqrt[4]{4\pi m}\left(\frac{2m}{\mathrm{e}}\right)^m} \right)^{2/m} \\
&= \lim_{m\to\infty} \frac{\left(\frac{L+2m-1}{\mathrm{e}}\right)^4}{4\left(\frac{L/2+m-1}{\mathrm{e}}\right)^2\left(\frac{2m}{\mathrm{e}}\right)^2} = 1.
\end{aligned}$$

As a conclusion, the stochastic process defined by (2.3)–(2.5) is an $\mathrm{L}^p(\Omega)$ solution to (2.1) on $(-1,1)$.
To demonstrate the uniqueness, we use [160, Th. 5.1.2], [161, Th. 5]. Rewrite (2.1) as $\dot{Z}(t) = B(t)Z(t)$, where

$$Z(t) = \begin{pmatrix} X(t) \\ \dot{X}(t) \end{pmatrix}, \quad B(t) = \begin{pmatrix} 0 & 1 \\ \frac{A(A+1)}{1-t^2} & \frac{-2t}{1-t^2} \end{pmatrix}.$$

We say that the random vector $Z = (Z_1, Z_2)$ belongs to $\mathrm{L}_2^p(\Omega)$ if

$$\|Z\|_p \coloneqq \max\{\|Z_1\|_p, \|Z_2\|_p\} < \infty.$$

Consider the random matrix norm $|||B||| \coloneqq \max_i \sum_j \|b_{ij}\|_\infty$. If $Z, Z' \in \mathrm{L}_2^p(\Omega)$, then $\|B(t)Z - B(t)Z'\|_p \leq |||B(t)||| \cdot \|Z - Z'\|_p$, where

$$\int_{-a}^{a} |||B(t)||| \, \mathrm{d}t = \int_{-a}^{a} \frac{\|A\|_\infty(\|A\|_\infty + 1) + 2|t|}{1 - t^2} \, \mathrm{d}t < \infty$$

for each $a \in (0,1)$. Then the assumptions of [160, Th. 5.1.2], [161, Th. 5] hold. □

The hypothesis $\|A\|_\infty < \infty$ is satisfied by some standard probability distributions: Uniform, Beta, Binomial, etc. If one wants A to follow an unbounded distribution, the truncation method permits bounding the support of A (see [125]). For example, the truncated Normal or Gamma distributions can be given to A. See Example 2.9 for a test of this methodology.

Remark 2.5 If $\|A\|_\infty = \infty$, then (2.6) does not hold for any $t \in (-1,1)\backslash\{0\}$. Indeed,

$$\sum_{m=0}^{\infty} \frac{1}{(2m)!} \|P_1(m)\|_\infty |t|^{2m} \geq \frac{1}{2}\|P_1(1)\|_\infty t^2 = \frac{1}{2}\|A(A+1)\|_\infty t^2 = \infty.$$

By Lemma 2.3, the two series given in (2.4) do not converge in $\mathrm{L}^\infty(\Omega)$, for any $t \in (-1,1)\backslash\{0\}$.

Remark 2.6 If $X_0, X_1, A \in \mathrm{L}^\infty(\Omega)$, then the response process $X(t)$ defined by (2.3)–(2.5) is the unique $\mathrm{L}^\infty(\Omega)$ solution to (2.1) on $(-1,1)$. In particular, $X(t)$ is the unique solution in the sample-path sense [160, Appendix A].

2.3 Approximation of the expectation and variance of the response process

Let $X_0, X_1 \in \mathrm{L}^2(\Omega)$ and A be a bounded random variable, not necessarily independent. By Theorem 2.4, the stochastic process $X(t)$ defined by (2.3)–(2.5) is an $\mathrm{L}^2(\Omega)$ solution to the random initial value problem (2.1) on the whole time domain $(-1,1)$. If we consider $X^M(t) = X_0\tilde{X}_1^M(t) + X_1\tilde{X}_2^M(t)$, where

$$\tilde{X}_1^M(t) = \sum_{m=0}^{\lfloor \frac{M}{2} \rfloor} \frac{(-1)^m}{(2m)!} P_1(m) t^{2m}, \quad \tilde{X}_2^M(t) = \sum_{m=0}^{\lfloor \frac{M-1}{2} \rfloor} \frac{(-1)^m}{(2m+1)!} P_2(m) t^{2m+1},$$

we know that $X^M(t) \to X(t)$ in $\mathrm{L}^2(\Omega)$ as $M \to \infty$, for each $t \in (-1,1)$. This mean square convergence allows us to approximate the expectation and variance of $X(t)$ by using

$$\mathbb{E}[X(t)] = \lim_{M\to\infty} \mathbb{E}[X^M(t)], \quad \mathbb{V}[X(t)] = \lim_{M\to\infty} \mathbb{V}[X^M(t)], \tag{2.7}$$

see [160, Th. 4.2.1, Th. 4.3.1].

The expectation of $X^M(t)$ is given by

$$\mathbb{E}[X^M(t)] = \sum_{m=0}^{\lfloor \frac{M}{2} \rfloor} \frac{(-1)^m}{(2m)!} \mathbb{E}[X_0 P_1(m)] t^{2m} + \sum_{m=0}^{\lfloor \frac{M-1}{2} \rfloor} \frac{(-1)^m}{(2m+1)!} \mathbb{E}[X_1 P_2(m)] t^{2m+1},$$

where

$$\mathbb{E}[X_0 P_1(m)] = \int_{(0,\infty)\times\mathbb{R}} x_0 \left(\prod_{j=1}^{m} (a-2j+2)(a+2j-1) \right) \mathbb{P}_{(A,X_0)}(\mathrm{d}a, \mathrm{d}x_0),$$

$$\mathbb{E}[X_1 P_2(m)] = \int_{(0,\infty)\times\mathbb{R}} x_1 \left(\prod_{j=1}^{m} (a-2j+1)(a+2j) \right) \mathbb{P}_{(A,X_1)}(\mathrm{d}a, \mathrm{d}x_1).$$

Here, $\mathbb{P}_Z$ represents the probability law of the random vector Z, which comprises the different cases of absolute continuity, discrete support, etc.

On the other hand, the variance of $X_M(t)$ is given by

$$\mathbb{V}[X_M(t)] = \mathbb{E}[X_M(t)^2] - \left(\mathbb{E}[X_M(t)]\right)^2,$$

so that we need to compute $\mathbb{E}[X_M(t)^2]$. Let

$$X_{2m} = X_0 \frac{(-1)^m}{(2m)!} P_1(m), \quad X_{2m+1} = X_1 \frac{(-1)^m}{(2m+1)!} P_2(m).$$

We have

$$\mathbb{E}[X_M(t)^2] = \mathbb{E}\left[\left(\sum_{m=0}^{\lfloor \frac{M}{2} \rfloor} X_{2m} t^{2m}\right)^2\right] + \mathbb{E}\left[\left(\sum_{m=0}^{\lfloor \frac{M-1}{2} \rfloor} X_{2m+1} t^{2m+1}\right)^2\right] + 2\sum_{m=0}^{\lfloor \frac{M}{2} \rfloor}\sum_{n=0}^{\lfloor \frac{M-1}{2} \rfloor} \mathbb{E}[X_{2m}X_{2n+1}] t^{2(m+n)+1},$$

where

$$\mathbb{E}\left[\left(\sum_{m=0}^{\lfloor \frac{M}{2} \rfloor} X_{2m} t^{2m}\right)^2\right] = \sum_{m=0}^{\lfloor \frac{M}{2} \rfloor}\sum_{n=0}^{\lfloor \frac{M}{2} \rfloor} \mathbb{E}[X_{2m}X_{2n}] t^{2(m+n)},$$

$$\mathbb{E}\left[\left(\sum_{m=0}^{\lfloor \frac{M-1}{2} \rfloor} X_{2m+1} t^{2m+1}\right)^2\right] = \sum_{m=0}^{\lfloor \frac{M-1}{2} \rfloor}\sum_{n=0}^{\lfloor \frac{M-1}{2} \rfloor} \mathbb{E}[X_{2m+1}X_{2n+1}] t^{2(m+n)+2}.$$

The expectations involved in these expressions can be computed as follows:

$$\begin{aligned}
\mathbb{E}[X_{2m}X_{2n}] &= \frac{(-1)^{m+n}}{(2m)!(2n)!}\mathbb{E}[X_0^2 P_1(m)P_1(n)] \\
&= \frac{(-1)^{m+n}}{(2m)!(2n)!}\int_{(0,\infty)\times\mathbb{R}} x_0^2 \left(\prod_{j=1}^{m}(a-2j+2)(a+2j-1)\right) \\
&\quad \cdot \left(\prod_{j=1}^{n}(a-2j+2)(a+2j-1)\right) \mathbb{P}_{(A,X_0)}(\mathrm{d}a, \mathrm{d}x_0),
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[X_{2m+1}X_{2n+1}] &= \frac{(-1)^{m+n}}{(2m+1)!(2n+1)!}\mathbb{E}[X_1^2 P_2(m)P_2(n)] \\
&= \frac{(-1)^{m+n}}{(2m+1)!(2n+1)!}\int_{(0,\infty)\times\mathbb{R}} x_1^2 \left(\prod_{j=1}^{m}(a-2j+1)(a+2j)\right) \\
&\quad \cdot \left(\prod_{j=1}^{n}(a-2j+1)(a+2j)\right) \mathbb{P}_{(A,X_1)}(\mathrm{d}a, \mathrm{d}x_1),
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[X_{2m}X_{2n+1}] &= \frac{(-1)^{m+n}}{(2m)!(2n+1)!}\mathbb{E}[X_0 X_1 P_1(m)P_2(n)] \\
&= \frac{(-1)^{m+n}}{(2m)!(2n)!}\int_{(0,\infty)\times\mathbb{R}\times\mathbb{R}} x_0 x_1 \left(\prod_{j=1}^{m}(a-2j+2)(a+2j-1)\right) \\
&\quad \cdot \left(\prod_{j=1}^{n}(a-2j+1)(a+2j)\right) \mathbb{P}_{(A,X_0,X_1)}(\mathrm{d}a, \mathrm{d}x_0, \mathrm{d}x_1).
\end{aligned}$$

2.4 Numerical experiments

In this section we perform several numerical experiments. Since in [31] the authors carried out numerical examples when A, X_0 and X_1 are independent random variables, we will show three more examples in which A, X_0 and X_1 are not independent. To assess the reliability of the approximations obtained for the expectation and variance by using (2.7), we will compare them with Monte Carlo simulation.

Monte Carlo simulation generates samples of $X(t)$ by computing realizations of A, X_0 and X_1 and solving the corresponding deterministic problem (2.1). Although it is an effective and easy to implement approach to quantify the uncertainty, the slowness to get accurately the digits in the computations makes this technique computationally expensive, [70], [173, pp. 53–54].

Example 2.7 We consider the random differential equation (2.1) with

$$(A, X_0, X_1) \sim \text{Dirichlet}(5, 1, 2, 3).$$

Since X_0, X_1 and A are bounded random variables, Theorem 2.4 implies that the stochastic process $X(t)$ defined by (2.3)–(2.5) is the unique $\mathrm{L}^\infty(\Omega)$ solution to (2.1) on $(-1, 1)$. In Table 2.1, we show $\mathbb{E}[X^M(t)]$ for different orders M, which approximates $\mathbb{E}[X(t)]$ by (2.7). We observe that the approximations achieved are more accurate for small M when t is near 0, because the random power series is centered at 0 and the process $X(t)$ is known at 0. For $t \leq 0.8$, stabilization of the results has been achieved for $M = 80$. For $t = 0.9$, a larger M would be needed. We notice that Monte Carlo simulation with $500,000$ realizations give an approximate result up to three significant figures. To obtain more exact approximations, more simulations and computational cost are needed. In general, the approximations via Monte Carlo simulation are worse than via our Fröbenius method. Table 2.2 provides analogous results for the variance, where $\mathbb{V}[X^M(t)]$ approximates $\mathbb{V}[X(t)]$ by (2.7). For $t \leq 0.7$ stabilization of the approximations has been reached for $M = 80$. In general, a larger M is required to achieve nearly exact approximations for the variance. The results obtained agree with the Monte Carlo simulation.

Example 2.8 We set a joint discrete distribution to (A, X_0, X_1):

$$(A, X_0, X_1) \sim \text{Multinomial}(10; 0.2, 0.3, 0.5).$$

Since X_0, X_1 and A are bounded random variables, Theorem 2.4 entails that the response process $X(t)$ defined by (2.3)–(2.5) is the unique $\mathrm{L}^\infty(\Omega)$ solution

t	$\mathbb{E}[X^{10}(t)]$	$\mathbb{E}[X^{20}(t)]$	$\mathbb{E}[X^{40}(t)]$	$\mathbb{E}[X^{80}(t)]$	MC $500,000$
0	0.0909091	0.0909091	0.0909091	0.0909091	0.0906787
0.1	0.108855	0.108855	0.108855	0.108855	0.108648
0.2	0.126491	0.126491	0.126491	0.126491	0.126308
0.3	0.144059	0.144059	0.144059	0.144059	0.143903
0.4	0.161835	0.161835	0.161835	0.161835	0.161709
0.5	0.180166	0.180172	0.180172	0.180172	0.180080
0.6	0.199548	0.199591	0.199592	0.199592	0.199540
0.7	0.220733	0.220998	0.221002	0.221002	0.221000
0.8	0.24491	0.246266	0.246352	0.246352	0.246416
0.9	0.273962	0.280100	0.281585	0.281693	0.281863

Table 2.1: Approximation of the expectation of the solution stochastic process. Example 2.7.

t	$\mathbb{V}[X^{10}(t)]$	$\mathbb{V}[X^{20}(t)]$	$\mathbb{V}[X^{40}(t)]$	$\mathbb{V}[X^{80}(t)]$	MC $500,000$
0	0.00688705	0.00688705	0.00688705	0.00688705	0.00685105
0.1	0.00670461	0.00670461	0.00670461	0.00670461	0.00666882
0.2	0.00672130	0.00672130	0.00672130	0.00672130	0.00668621
0.3	0.00697044	0.00697045	0.00697045	0.00697045	0.00693658
0.4	0.00751088	0.00751091	0.00751091	0.00751091	0.00747887
0.5	0.00844437	0.00844482	0.00844482	0.00844482	0.00841536
0.6	0.00995308	0.00995823	0.00995825	0.00995825	0.00993237
0.7	0.0123829	0.0124269	0.0124276	0.0124276	0.0124068
0.8	0.0164346	0.0167508	0.0167712	0.0167714	0.0167582
0.9	0.0236175	0.0256974	0.0262304	0.0262699	0.0262712

Table 2.2: Approximation of the variance of the solution stochastic process. Example 2.7.

t	$\mathbb{E}[X^{10}(t)]$	$\mathbb{E}[X^{20}(t)]$	$\mathbb{E}[X^{40}(t)]$	$\mathbb{E}[X^{80}(t)]$	MC $500,000$
0	3	3	3	3	3.00207
0.1	3.39965	3.39965	3.39965	3.39965	3.40154
0.2	3.59067	3.59067	3.59067	3.59067	3.59226
0.3	3.57194	3.57194	3.57194	3.57194	3.57322
0.4	3.35661	3.35661	3.35661	3.35661	3.35768
0.5	2.97154	2.97154	2.97154	2.97154	2.97259
0.6	2.45625	2.45623	2.45623	2.45623	2.45738
0.7	1.86122	1.86112	1.86111	1.86111	1.86226
0.8	1.24584	1.24523	1.24515	1.24515	1.24558
0.9	0.675881	0.672543	0.670722	0.670550	0.668041

Table 2.3: Approximation of the expectation of the solution stochastic process. Example 2.8.

to (2.1) on $(-1,1)$. Expression (2.7) allows approximating $\mathbb{E}[X(t)]$ and $\mathbb{V}[X(t)]$ via $\mathbb{E}[X^M(t)]$ and $\mathbb{V}[X^M(t)]$, respectively. Analogous comments to the previous example apply here, and the results are presented in Table 2.3 and Table 2.4. The results obtained via our Fröbenius method are accurate.

Example 2.9 We set a truncated Multinormal distribution for the random input parameters:

$$(A, X_0, X_1) \sim \text{Multinormal}\left(\begin{pmatrix}10\\-2\\1\end{pmatrix}, \begin{pmatrix}1 & 0.01 & -0.02\\0.01 & 4 & 2\\-0.02 & 2 & 4\end{pmatrix}\right)\Big|_{[6,14]\times\mathbb{R}\times\mathbb{R}}.$$

Since $X_0, X_1 \in \mathrm{L}^p(\Omega)$ for all $1 \leq p < \infty$ and A is bounded in $[6,14]$, Theorem 2.4 shows that the stochastic process $X(t)$ defined by (2.3)–(2.5) is the unique $\mathrm{L}^p(\Omega)$ solution to (2.1) on $(-1,1)$, for each $1 \leq p < \infty$. Analogously to the previous two examples, Table 2.5 and Table 2.6 show the results. Observe that stabilization of the results for $t \leq 0.7$ is achieved for $M = 80$. Notice also that, for $M \leq 20$ and $t \geq 0.4$, the approximation of the expectation and variance is not good. The results obtained from the Fröbenius method for $M \geq 80$ agree with the statistics calculated via the Monte Carlo simulation.

t	$\mathbb{V}[X^{10}(t)]$	$\mathbb{V}[X^{20}(t)]$	$\mathbb{V}[X^{40}(t)]$	$\mathbb{V}[X^{80}(t)]$	MC $500,000$
0	2.1	2.1	2.1	2.1	2.10094
0.1	1.81331	1.81331	1.81331	1.81331	1.81300
0.2	1.78089	1.78089	1.78089	1.78089	1.77973
0.3	2.43304	2.43304	2.43304	2.43304	2.43226
0.4	4.15996	4.15996	4.15996	4.15996	4.16027
0.5	7.12080	7.12077	7.12077	7.12077	7.12084
0.6	11.1844	11.1838	11.1838	11.1838	11.1812
0.7	16.1150	16.1090	16.1090	16.1090	16.1046
0.8	22.1109	22.0920	22.0932	22.0932	22.0965
0.9	31.1044	31.4557	31.6189	31.6324	31.6569

Table 2.4: Approximation of the variance of the solution stochastic process. Example 2.8.

t	$\mathbb{E}[X^{10}(t)]$	$\mathbb{E}[X^{20}(t)]$	$\mathbb{E}[X^{40}(t)]$	$\mathbb{E}[X^{80}(t)]$	MC $500,000$
0	-2.01642	-2.01642	-2.01642	-2.01642	-2.00100
0.1	-0.905676	-0.905676	-0.905676	-0.905676	-0.905209
0.2	1.10885	1.10884	1.10884	1.10884	1.11031
0.3	1.94955	1.94909	1.94909	1.94909	1.94966
0.4	0.656784	0.643176	0.643176	0.643176	0.641893
0.5	-1.20831	-1.39804	-1.39804	-1.39804	-1.39941
0.6	0.111123	-1.57901	-1.57903	-1.57903	-1.57838
0.7	10.8410	0.602665	0.602084	0.602084	0.594087
0.8	51.7915	1.58890	1.57617	1.57615	1.57588
0.9	203.700	-0.987211	-1.20468	-1.20776	-1.20091

Table 2.5: Approximation of the expectation of the solution stochastic process. Example 2.9.

t	$\mathbb{V}[X^{10}(t)]$	$\mathbb{V}[X^{20}(t)]$	$\mathbb{V}[X^{40}(t)]$	$\mathbb{V}[X^{80}(t)]$	MC $500,000$
0	3.96931	3.96931	3.96931	3.96931	4.00268
0.1	1.23016	1.23016	1.23016	1.23016	1.22715
0.2	1.16167	1.16166	1.16166	1.16166	1.14804
0.3	3.86797	3.87079	3.87079	3.87079	3.86348
0.4	1.72091	1.76984	1.76984	1.76984	1.76343
0.5	2.59759	2.75802	2.75802	2.75802	2.71796
0.6	53.8179	3.79667	3.79665	3.79665	3.79030
0.7	1774.74	3.88379	3.87941	3.87941	3.88103
0.8	40373.8	5.25517	5.27273	5.27282	5.17336
0.9	658630	4.79558	7.67724	7.76726	7.73295

Table 2.6: Approximation of the variance of the solution stochastic process. Example 2.9.

2.5 Conclusions

In this chapter we have studied the random Legendre differential equation with input coefficient A and initial conditions X_0 and X_1. In [Calbo G. et al, Comput. Math. Appl., 61(9), 2782–2792 (2011)], a mean square convergent random power series solution $X(t)$ on $(-1/\mathrm{e}, 1/\mathrm{e})$ was constructed via the Fröbenius method. The authors proved that, under the assumption that the absolute moments of A grow at most exponentially, under mean fourth integrability of X_0 and X_1, and under independence of A and the initial conditions, the random power series becomes a mean square solution to the random Legendre differential equation on $(-1/\mathrm{e}, 1/\mathrm{e})$. We have extended this result by assuming less integrability of X_0 and X_1 and no independence between the random inputs. Moreover, the growth condition on the absolute moments of A has been characterized in terms of the boundedness of A. This has permitted a simpler proof of our result, as no probabilistic inequalities (Hölder, c_s, etc.) have been required. Moreover, our random power series solution converges on the whole $(-1, 1)$, as it occurs with its deterministic counterpart. We have provided expressions for the approximate expectation and variance of $X(t)$, by truncating the random power series. In the numerical examples, we have illustrated the improvements developed by working with non-independent random inputs. Our approach has improved the approximations from the Monte Carlo simulation.

Acknowledgements

This work has been supported by the Spanish Ministerio de Economía y Competitividad grant MTM2017–89664–P. I acknowledge the doctorate scholarship granted by Programa de Ayudas de Investigación y Desarrollo (PAID), Universitat Politècnica de València.

The main results of this chapter have been published in [21].

Chapter 3

Random non-autonomous second-order linear differential equations: mean square analytic solutions and their statistical properties

In this chapter we study random non-autonomous second-order linear differential equations. The coefficients are assumed to be stochastic processes and the initial conditions are random variables defined in a common underlying complete probability space. Under appropriate assumptions, we prove the existence of an analytic stochastic process solution in the random mean square sense. Truncating the random series that defines the solution process, we are able to approximate the main statistical properties of the solution, such as the expectation and the variance. We also obtain error priori bounds to construct reliable approximations of both statistical moments. We include a set of numerical examples to illustrate the main theoretical results established throughout the chapter. We finish with an example where our findings are combined with Monte Carlo simulation to modeling uncertainty using real data.

3.1 Introduction

In this chapter, we conduct a full probabilistic study of the random second-order linear differential equation

$$\begin{cases} \ddot{X}(t) + A(t)\dot{X}(t) + B(t)X(t) = 0, \ t \in \mathbb{R}, \\ X(t_0) = Y_0, \\ \dot{X}(t_0) = Y_1. \end{cases} \tag{3.1}$$

The data coefficients $A(t)$ and $B(t)$ are stochastic processes and the initial conditions Y_0 and Y_1 are random variables on an underlying complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Particular cases of the random initial value problem (3.1) have been studied in previous contributions using $\mathrm{L}^p(\Omega)$ random calculus. As briefly mentioned in the previous chapter, important deterministic models appearing in the area of Mathematical Physics, such as Airy, Hermite, Legendre, Laguerre and Bessel differential equations, have been randomized and rigorously studied in [29, 31, 43, 44, 46] and in Chapter 2. In these contributions, approximate solution stochastic processes together with their main statistical moments (mean and variance) are constructed by taking advantage of the random mean square calculus. Since in the case of Hermite, Legendre and Laguerre deterministic differential equations it is well-known that they admit polynomial solutions, in those contributions the concept of random polynomial solution is introduced in the stochastic framework as well.

In [77], the authors proposed a homotopy technique to solve some particular random differential equations pertaining to the class given in (3.1). Other solution techniques include variational iteration [99] and Adomian decomposition [100]. Finally, a technique, analogous to the Fröbenius method but relying on the concept of differential transform, is proposed in [101, 168].

A very important case of problem (3.1) is when its coefficients are random variables rather than stochastic processes, i.e., $A(t) = A$ and $B(t) = B$, corresponding to the autonomous case. In [38] the authors construct approximations of the first and second probability density function of the solution stochastic process using a complementary approach to mean square calculus. In [57, 59, 87, 89, 90, 143] one addresses significant advances, for other random differential equations, dealing with the computation of the probability density function of the corresponding solution. Additional studies dealing with random differential equations via random mean square calculus include [97, 98, 100, 101, 115, 127], for instance.

The structure of this chapter is described as follows. In Section 3.2 we will solve the random initial value problem (3.1) in a suitable way and we will describe the manner of approximating the main statistical information of the solution process (mean and variance). In Section 3.3 we will compare our findings with the extant literature and we will also perform some numerical examples including an illustrative application in a modeling setting using real data. Finally, in Section 3.4 conclusions will be drawn.

3.2 Results

Our main goal is to find the solution stochastic process to the random initial value problem (3.1). We will assume that the data stochastic process $A(t)$ and $B(t)$ are analytic at t_0, in the following sense:

$$A(t) = \sum_{n=0}^{\infty} A_n (t - t_0)^n, \quad B(t) = \sum_{n=0}^{\infty} B_n (t - t_0)^n,$$

for $t \in (t_0 - r, t_0 + r)$, being $r > 0$ fixed, and the sum is understood in the $\mathrm{L}^2(\Omega)$ setting. We search for an analytic solution process $X(t)$ of the form

$$X(t) = \sum_{n=0}^{\infty} X_n (t - t_0)^n,$$

for $t \in (t_0 - r, t_0 + r)$, where the sum is in $\mathrm{L}^2(\Omega)$. This stochastic process will be a solution to the random problem (3.1) in the sense of $\mathrm{L}^2(\Omega)$ (so, in particular twice differentiable in the mean square sense).

3.2.1 Auxiliary results concerning random power series

We need some auxiliary results to deal with random power series in the $\mathrm{L}^2(\Omega)$ setting. First of all, we need a result to differentiate a power series in the $\mathrm{L}^p(\Omega)$ sense (in this chapter we will just use the cases $p = 1$ and $p = 2$, but we do the proof for a general $p \geq 1$ just for the sake of completeness). The particular case $p = 2$ is a consequence of Theorem 3.1 in [52].

Theorem 3.1 (Differentiation of a power series in the $\mathrm{L}^p(\Omega)$ sense) *Let $A(t) = \sum_{n=0}^{\infty} A_n (t - t_0)^n$ be a random power series in the $\mathrm{L}^p(\Omega)$ setting ($p \geq 1$), for $t \in (t_0 - r, t_0 + r)$, $r > 0$. Then the random power series $\sum_{n=1}^{\infty} n A_n (t - t_0)^{n-1}$ exists in $\mathrm{L}^p(\Omega)$ for $t \in (t_0 - r, t_0 + r)$ and, moreover, the*

$\mathrm{L}^p(\Omega)$ *derivative of* $A(t)$ *is equal to it:*

$$\dot{A}(t) = \sum_{n=1}^{\infty} nA_n(t-t_0)^{n-1},$$

for all $t \in (t_0 - r, t_0 + r)$.

Proof. Let us see first that the random power series $\sum_{n=1}^{\infty} nA_n(t-t_0)^{n-1}$ exists in $\mathrm{L}^p(\Omega)$ for $t \in (t_0 - r, t_0 + r)$. Given $0 < \rho < r$, we prove that

$$\sum_{n=1}^{\infty} n\|A_n\|_p \rho^{n-1} < \infty. \tag{3.2}$$

Fix $s : 0 < \rho < s < r$. Since the sum $\sum_{n=0}^{\infty} A_n s^n$ exists in $\mathrm{L}^p(\Omega)$, we have $\lim_{n\to\infty} \|A_n\|_p s^n = 0$, so there exists $K > 0$ such that $\|A_n\|_p s^n \leq K$, for every $n \geq 0$. Then

$$n\|A_n\|_p \rho^{n-1} \leq n\,K\,\frac{1}{s}\left(\frac{\rho}{s}\right)^{n-1}.$$

As $\sum_{n=1}^{\infty} n(\rho/s)^{n-1} < \infty$, by the comparison test for series we conclude that (3.2) holds.
Let us see now that the $\mathrm{L}^p(\Omega)$ derivative of $A(t)$ is equal to $\sum_{n=1}^{\infty} nA_n(t-t_0)^{n-1}$. By definition of $\mathrm{L}^p(\Omega)$ derivative, we have to check that

$$\lim_{h\to 0} \left\| \sum_{n=0}^{\infty} A_n \frac{(t+h-t_0)^n - (t-t_0)^n}{h} - \sum_{n=1}^{\infty} nA_n(t-t_0)^{n-1} \right\|_p = 0.$$

Fix s and h such that $0 < \rho < s < r$ and $0 < |h| < s - \rho$. Fix $t \in (t_0 - \rho, t_0 + \rho)$. By the triangular inequality,

$$\begin{aligned} &\left\| \sum_{n=0}^{\infty} A_n \frac{(t+h-t_0)^n - (t-t_0)^n}{h} - \sum_{n=1}^{\infty} nA_n(t-t_0)^{n-1} \right\|_p \\ &\leq \sum_{n=1}^{\infty} \|A_n\|_p \left| \frac{(t+h-t_0)^n - (t-t_0)^n}{h} - n(t-t_0)^{n-1} \right|. \end{aligned} \tag{3.3}$$

We know that

$$\lim_{h\to 0} \left| \frac{(t+h-t_0)^n - (t-t_0)^n}{h} - n(t-t_0)^{n-1} \right| = 0 \tag{3.4}$$

(by definition of the pointwise derivative of $(t-t_0)^n$). On the other hand, using the identity $a^n - b^n = (a-b)(\sum_{m=0}^{n-1} a^{n-1-m}b^m)$, we perform the following

estimates:

$$\begin{aligned}\left|\frac{(t+h-t_0)^n-(t-t_0)^n}{h}\right| &= \left|\frac{(t+h-t_0)^n-(t-t_0)^n}{(t+h-t_0)-(t-t_0)}\right| \\ &= \left|\sum_{m=0}^{n-1}(t+h-t_0)^{n-1-m}(t-t_0)^m\right| \\ &\leq \sum_{m=0}^{n-1}|t+h-t_0|^{n-1-m}|t-t_0|^m \\ &\leq \sum_{m=0}^{n-1}(|t-t_0|+|h|)^{n-1-m}|t-t_0|^m \\ &\leq n(|t-t_0|+|h|)^{n-1} \leq ns^{n-1},\end{aligned}$$

where we have used that $|t-t_0| \leq |t-t_0|+|h|$, $|t-t_0|<\rho$ and $|h|<s-\rho$. Then

$$\begin{aligned}&\|A_n\|_p\left|\frac{(t-t_0+h)^n-(t-t_0)^n}{h}-n(t-t_0)^{n-1}\right| \\ &\leq \|A_n\|_p\left(\left|\frac{(t-t_0+h)^n-(t-t_0)^n}{h}\right|+n|t-t_0|^{n-1}\right) \\ &\leq \|A_n\|_p(ns^{n-1}+n\rho^{n-1}) \leq \|A_n\|_p(ns^{n-1}+ns^{n-1}) = 2\|A_n\|_p ns^{n-1}, \qquad (3.5)\end{aligned}$$

being $\sum_{n=1}^{\infty}\|A_n\|_p ns^{n-1}<\infty$. By the Dominated Convergence Theorem for series, both (3.4) and (3.5) permit us to conclude that (3.3) tends to 0 as $h\to 0$, as wanted.

□

As it shall see apparent later, we need a theorem to multiply random power series. In the deterministic setting, the so-called Merten's Theorem allows multiplying two power series. The deterministic version of Merten's Theorem is proved in Theorem 8.46 of [3]. We adapt the proof in [3] to a stochastic setting in Theorem 3.2. Notice that, from Theorem 3.2, when we multiply two random power series we loose Lebesgue spaces of convergence. As we will see in the proof, this fact will be a consequence of Cauchy-Schwarz inequality (for two real numbers u and v, we always have $|uv|=|u||v|$, however, for random variables U and V we do not have $\|UV\|_2 = \|U\|_2\|V\|_2$ in general, but $\|UV\|_1 \leq \|U\|_2\|V\|_2$).

Theorem 3.2 (Merten's Theorem in the mean square sense) *Let* $U=\sum_{n=0}^{\infty}U_n$ *and* $V=\sum_{n=0}^{\infty}V_n$ *be two random series that converge in* $\mathrm{L}^2(\Omega)$. *Sup-*

pose one of the series converges absolutely, say $\sum_{n=0}^{\infty} \|V_n\|_2 < \infty$. *Then*

$$\left(\sum_{n=0}^{\infty} U_n\right)\left(\sum_{n=0}^{\infty} V_n\right) = \sum_{n=0}^{\infty} W_n,$$

where

$$W_n = \sum_{m=0}^{n} U_{n-m} V_m$$

and $\sum_{n=0}^{\infty} W_n$ *is understood in* $\mathrm{L}^1(\Omega)$. *The series* $\sum_{n=0}^{\infty} W_n$ *is known as the Cauchy product of the series* $\sum_{n=0}^{\infty} U_n$ *and* $\sum_{n=0}^{\infty} V_n$.

Proof. Let us write the N-th partial sum of $\sum_{n=0}^{\infty} W_n$ in an appropriate way:

$$\begin{aligned}\sum_{n=0}^{N} W_n &= \sum_{n=0}^{N}\sum_{m=0}^{n} U_{n-m}V_m = \sum_{m=0}^{N} V_m \sum_{n=m}^{N} U_{n-m} = \sum_{m=0}^{N} V_m \sum_{n=0}^{N-m} U_n \\ &= \sum_{m=0}^{N} V_m \left(U - \sum_{n=N-m+1}^{\infty} U_n\right) = U\sum_{m=0}^{N} V_m - \sum_{m=0}^{N} V_m \sum_{n=N-m+1}^{\infty} U_n.\end{aligned}$$

The first addend, $U\sum_{m=0}^{N} V_m$, tends to UV in $\mathrm{L}^1(\Omega)$ as $N \to \infty$. In fact, observe that because of Cauchy-Schwarz inequality, one gets

$$\left\| U\sum_{m=0}^{N} V_m - UV \right\|_1 = \left\| U\left(\sum_{m=0}^{N} V_m - V\right)\right\|_1 \leq \|U\|_2 \left\|\sum_{m=0}^{N} V_m - V\right\|_2 \xrightarrow[N\to\infty]{} 0,$$

where we have used that $\|U\|_2 < \infty$ (since $U \in \mathrm{L}^2(\Omega)$ because it is the m.s. limit of the series $\sum_{n=0}^{\infty} U_n$) and that $\left\|\sum_{m=0}^{N} V_m - V\right\|_2 \xrightarrow[N\to\infty]{} 0$ (since by hypothesis $\sum_{n=0}^{\infty} V_n$ converges to V in $\mathrm{L}^2(\Omega)$).
Thus, it remains to prove that the second addend, $\sum_{m=0}^{N} V_m \sum_{n=N-m+1}^{\infty} U_n$, goes to 0 in $\mathrm{L}^1(\Omega)$ as $N \to \infty$.
Since $\lim_{N\to\infty} \sum_{n=N}^{\infty} U_n = 0$ in $\mathrm{L}^2(\Omega)$, there exists $L > 0$ such that

$$\left\|\sum_{n=N}^{\infty} U_n\right\|_2 \leq L, \quad \forall N \in \mathbb{N}.$$

Let $K = \sum_{n=0}^{\infty} \|V_n\|_2$. Fix $\epsilon > 0$. We can take N_ϵ such that, for all $N \geq N_\epsilon$,

$$\left\|\sum_{n=N}^{\infty} U_n\right\|_2 < \frac{\epsilon}{2K}, \quad \sum_{n=N+1}^{\infty} \|V_n\|_2 < \frac{\epsilon}{2L}.$$

Then, for $N \geq 2N_\epsilon$, by the triangular inequality and Cauchy-Schwarz inequality,

$$\begin{aligned}\left\|\sum_{m=0}^{N} V_m \sum_{n=N-m+1}^{\infty} U_n\right\|_1 &\leq \sum_{m=0}^{N_\epsilon} \|V_m\|_2 \left\|\sum_{n=N-m+1}^{\infty} U_n\right\|_2 \\ &\quad + \sum_{m=N_\epsilon+1}^{N} \|V_m\|_2 \left\|\sum_{n=N-m+1}^{\infty} U_n\right\|_2 \\ &\leq \frac{\epsilon}{2K} \sum_{m=0}^{N_\epsilon} \|V_m\|_2 + L \sum_{m=N_\epsilon+1}^{N} \|V_m\|_2 \\ &\leq \frac{\epsilon}{2K} \sum_{m=0}^{\infty} \|V_m\|_2 + L \sum_{m=N_\epsilon+1}^{\infty} \|V_m\|_2 \\ &\leq \frac{\epsilon}{2K} K + L\frac{\epsilon}{2L} = \epsilon.\end{aligned}$$

This shows that $\sum_{m=0}^{N} V_m \sum_{n=N-m+1}^{\infty} U_n$ tends to 0 in $\mathrm{L}^1(\Omega)$ as $N \to \infty$.

□

3.2.2 Main result: Constructing the solution stochastic process to the random non-autonomous second-order linear differential equation

We present the main results of this chapter. After stating and proving them, a deeper analysis of the hypotheses and consequences will be performed.

Theorem 3.3 *Let $A(t) = \sum_{n=0}^{\infty} A_n (t - t_0)^n$ and $B(t) = \sum_{n=0}^{\infty} B_n (t - t_0)^n$ be two random series in the $\mathrm{L}^2(\Omega)$ setting, for $t \in (t_0 - r, t_0 + r)$, being $r > 0$ finite and fixed. Assume that the initial conditions Y_0 and Y_1 belong to $\mathrm{L}^2(\Omega)$. Suppose that there is a constant $C_r > 0$, maybe dependent on r, such that $\|A_n\|_\infty \leq C_r/r^n$ and $\|B_n\|_\infty \leq C_r/r^n$, $n \geq 0$. Then the stochastic process $X(t) = \sum_{n=0}^{\infty} X_n (t - t_0)^n$, $t \in (t_0 - r, t_0 + r)$, where*

$$X_0 = Y_0, \quad X_1 = Y_1, \tag{3.6}$$

$$X_{n+2} = \frac{-1}{(n+2)(n+1)} \sum_{m=0}^{n} \left[(m+1) A_{n-m} X_{m+1} + B_{n-m} X_m\right], \; n \geq 0, \tag{3.7}$$

is the unique analytic solution to the random initial value problem (3.1) in the mean square sense.

Proof. Suppose that $X(t) = \sum_{n=0}^{\infty} X_n (t-t_0)^n$ is a solution to (3.1) in the $\mathrm{L}^2(\Omega)$ sense, for $t \in (t_0 - r, t_0 + r)$, $r > 0$. By Theorem 3.1 with $p = 2$, the mean square derivatives of $X(t)$ are given by:

$$\dot{X}(t) = \sum_{n=1}^{\infty} n X_n (t-t_0)^{n-1} = \sum_{n=0}^{\infty} (n+1) X_{n+1} (t-t_0)^n,$$

$$\ddot{X}(t) = \sum_{n=2}^{\infty} n(n-1) X_n (t-t_0)^{n-2} = \sum_{n=0}^{\infty} (n+2)(n+1) X_{n+2} (t-t_0)^n.$$

By the random Merten's Theorem 3.2,

$$A(t)\dot{X}(t) = \sum_{n=0}^{\infty} \left(\sum_{m=0}^{n} A_{n-m} (m+1) X_{m+1} \right) (t-t_0)^n,$$

$$B(t)X(t) = \sum_{n=0}^{\infty} \left(\sum_{m=0}^{n} B_{n-m} X_m \right) (t-t_0)^n,$$

where these two random series converge in $\mathrm{L}^1(\Omega)$. From $\ddot{X}(t) + A(t)\dot{X}(t) + B(t)X(t) = 0$,

$$\sum_{n=0}^{\infty} \left[(n+2)(n+1) X_{n+2} + \sum_{m=0}^{n} \left(A_{n-m} (m+1) X_{m+1} + B_{n-m} X_m \right) \right] (t-t_0)^n = 0, \tag{3.8}$$

for $t \in (t_0 - r, t_0 + r)$, where the random series is again in the $\mathrm{L}^1(\Omega)$ sense. By Theorem 3.1 with $p = 1$, differentiating (3.8) over and over again in the $\mathrm{L}^1(\Omega)$ sense and evaluating at $t = t_0$ yield

$$(n+2)(n+1) X_{n+2} + \sum_{m=0}^{n} \left(A_{n-m} (m+1) X_{m+1} + B_{n-m} X_m \right) = 0.$$

Isolating X_{n+2} we obtain the recursive expression (3.7):

$$X_{n+2} = \frac{-1}{(n+2)(n+1)} \sum_{m=0}^{n} \left[(m+1) A_{n-m} X_{m+1} + B_{n-m} X_m \right].$$

The initial conditions of the random initial value problem (3.1) give (3.6), and so $X(t)$ is uniquely determined with probability 1.
It remains to check that the random series $\sum_{n=0}^{\infty} X_n (t-t_0)^n$ is convergent in $\mathrm{L}^2(\Omega)$. For that purpose, we will make use of the $\mathrm{L}^{\infty}(\Omega)$ bounds for A_n and B_n quoted in the hypotheses.

From the hypothesis $Y_0, Y_1 \in \mathrm{L}^2(\Omega)$ and by induction on n in expression (3.7), we obtain that $X_n \in \mathrm{L}^2(\Omega)$ for all $n \geq 0$. By the triangular inequality and the hypotheses,

$$\begin{aligned}\|X_{n+2}\|_2 &\leq \frac{1}{(n+2)(n+1)} \sum_{m=0}^{n} \left[(m+1)\|A_{n-m}X_{m+1}\|_2 + \|B_{n-m}X_m\|_2\right] \\ &\leq \frac{1}{(n+2)(n+1)} \frac{C_r}{r^n} \sum_{m=0}^{n} r^m \left((m+1)\|X_{m+1}\|_2 + \|X_m\|_2\right). \end{aligned} \tag{3.9}$$

Define $H_0 := \|Y_0\|_2$, $H_1 := \|Y_1\|_2$ and

$$H_{n+2} := \frac{1}{(n+2)(n+1)} \frac{C_r}{r^n} \sum_{m=0}^{n} r^m \left((m+1)H_{m+1} + H_m\right). \tag{3.10}$$

By induction on n it is trivially seen that $\|X_n\|_2 \leq H_n$, for $n \geq 0$. Thus, given $0 < \rho < r$, it is enough to see that $\sum_{n=0}^{\infty} H_n \rho^n < \infty$. For that purpose, we rewrite (3.10) so that H_{n+2} is expressed as a function of H_{n+1} and H_n (second-order recurrence equation):

$$\begin{aligned} H_{n+2} &= \frac{1}{(n+2)(n+1)} \frac{C_r}{r^n} \left(\sum_{m=0}^{n-1} r^m \left((m+1)H_{m+1} + H_m\right) + r^n \left((n+1)H_{n+1} + H_n\right) \right) \\ &= \frac{1}{(n+2)(n+1)} \frac{C_r}{r^n} \frac{(n+1)n}{C_r} r^{n-1} \Bigg(\underbrace{\frac{1}{(n+1)n} \frac{C_r}{r^{n-1}} \sum_{m=0}^{n-1} r^m \left((m+1)H_{m+1} + H_m\right)}_{=H_{n+1}} \Bigg) \\ &\quad + \frac{C_r}{n+2} H_{n+1} + \frac{C_r}{(n+2)(n+1)} H_n \\ &= \left(\frac{n}{(n+2)r} + \frac{C_r}{n+2} \right) H_{n+1} + \frac{C_r}{(n+2)(n+1)} H_n. \end{aligned} \tag{3.11}$$

Fix $s : 0 < \rho < s < r$. We have

$$H_{n+2} s^{n+2} = \left(\frac{ns}{(n+2)r} + \frac{C_r s}{n+2} \right) H_{n+1} s^{n+1} + \frac{C_r s^2}{(n+2)(n+1)} H_n s^n.$$

Let $M_n = \max_{0 \leq m \leq n} H_m s^m$. We have

$$H_{n+2} s^{n+2} \leq \left(\frac{ns}{(n+2)r} + \frac{C_r s}{n+2} + \frac{C_r s^2}{(n+2)(n+1)} \right) M_{n+1}. \tag{3.12}$$

Since

$$\lim_{n \to \infty} \frac{ns}{(n+2)r} + \frac{C_r s}{n+2} + \frac{C_r s^2}{(n+2)(n+1)} = \frac{s}{r} < 1,$$

it holds $M_{n+2} = M_{n+1}$ for all large n, and call the common value M. Hence, $H_n s^n \leq M$ for all large n, therefore $H_n \rho^n \leq M(\rho/s)^n$. Since $\sum_{n=0}^{\infty}(\rho/s)^n < \infty$, by comparison the series $\sum_{n=0}^{\infty} H_n \rho^n$ converges, and we are done.

□

Theorem 3.4 *Let $A(t) = \sum_{n=0}^{\infty} A_n(t-t_0)^n$ and $B(t) = \sum_{n=0}^{\infty} B_n(t-t_0)^n$ be two random series in the $\mathrm{L}^\infty(\Omega)$ setting, for $t \in (t_0 - r, t_0 + r)$, being $r > 0$ finite and fixed. Assume that the initial conditions Y_0 and Y_1 belong to $\mathrm{L}^2(\Omega)$. Then the stochastic process $X(t) = \sum_{n=0}^{\infty} X_n(t-t_0)^n$, $t \in (t_0 - r, t_0 + r)$, whose coefficients are defined by (3.6)–(3.7), is the unique analytic solution to the random initial value problem (3.1) in the mean square sense.*

Proof. By Lemma 2.3 in Chapter 2,

$$\sum_{n=0}^{\infty} \|A_n\|_\infty |t-t_0|^n < \infty, \quad \sum_{n=0}^{\infty} \|B_n\|_\infty |t-t_0|^n < \infty,$$

for $t \in (t_0 - r, t_0 + r)$. Thus, for each $0 \leq r_1 < r$,

$$\sum_{n=0}^{\infty} \|A_n\|_\infty r_1^n < \infty, \quad \sum_{n=0}^{\infty} \|B_n\|_\infty r_1^n < \infty.$$

Since the sequences $\{\|A_n\|_\infty r_1^n\}_{n=0}^{\infty}$ and $\{\|B_n\|_\infty r_1^n\}_{n=0}^{\infty}$ tend to 0, they are both bounded by a number $C_{r_1} > 0$:

$$\|A_n\|_\infty \leq \frac{C_{r_1}}{r_1^n}, \quad \|B_n\|_\infty \leq \frac{C_{r_1}}{r_1^n}, \quad n \geq 0.$$

Then Theorem 3.3 is applicable with r_1: the process $X(t) = \sum_{n=0}^{\infty} X_n(t-t_0)^n$ whose coefficients are given by (3.6)–(3.7) is a mean square solution to (3.1) on $(t_0 - r_1, t_0 + r_1)$. Now, since r_1 is arbitrary, we can can extend this result to the whole interval $(t_0 - r, r_0 + r)$.

□

To deal with uniqueness, we use an habitual extension of the classical Picard's theorem to mean square calculus [160, Th. 5.1.2].

Theorem 3.5 *If $A(t)$ and $B(t)$ are continuous stochastic processes in the $\mathrm{L}^\infty(\Omega)$ sense, then the mean square solution to (3.1) is unique.*

Proof. We write (3.1) as a first-order random differential equation, which is the setting under study in [160]:

$$\underbrace{\begin{pmatrix}\dot{X}(t)\\ \ddot{X}(t)\end{pmatrix}}_{\dot{Z}(t)} = \underbrace{\begin{pmatrix}0 & 1\\ -B(t) & -A(t)\end{pmatrix}}_{M(t)} \underbrace{\begin{pmatrix}X(t)\\ \dot{X}(t)\end{pmatrix}}_{Z(t)} + \underbrace{\begin{pmatrix}0\\ C(t)\end{pmatrix}}_{q(t)}.$$

We work in the space $\mathrm{L}_2^2(\Omega)$ of 2-dimensional random vectors whose components belong to $\mathrm{L}^2(\Omega)$. Given $Z = (Z_1, Z_2) \in \mathrm{L}_2^2(\Omega)$, its norm is defined as

$$\|Z\|_2 = \max\{\|Z_1\|_2, \|Z_2\|_2\}.$$

On the other hand, given a random matrix $B = (B_{ij})$, we define the following norm:

$$|||B||| = \max_i \sum_j \|B_{ij}\|_\infty.$$

In the case of the random matrix $M(t)$, it holds

$$|||M(t)||| = \max\{1, \|A(t)\|_\infty + \|B(t)\|_\infty\}. \tag{3.13}$$

Given $Z, Z' \in \mathrm{L}_2^2(\Omega)$, we have

$$\|(M(t)Z+q(t))-(M(t)Z'+q(t))\|_2 = \|M(t)(Z-Z')\|_2 \le \underbrace{|||M(t)|||}_{k(t)} \cdot \|Z-Z'\|_2.$$

Since $A(t)$ and $B(t)$ are continuous stochastic processes in the $\mathrm{L}^\infty(\Omega)$ sense, the real maps

$$t \in (t_0 - r, t_0 + r) \mapsto \|A(t)\|_\infty, \quad t \in (t_0 - r, t_0 + r) \mapsto \|B(t)\|_\infty$$

are continuous. By (3.13), the deterministic function $k(t)$ is continuous on $(t_0 - r, t_0 + r)$. This implies that $k \in \mathrm{L}^1([t_0 - r_1, t_0 + r_1])$ for each $0 < r_1 < r$. By [160, Th. 5.1.2], there is uniqueness of mean square solution for (3.1) on $[t_0 - r_1, t_0 + r_1]$. Since r_1 is arbitrary, there is uniqueness of solution on $(t_0 - r, t_0 + r)$.

□

3.2.3 Comments on the hypotheses of the theorems

The hypotheses concerning the $\mathrm{L}^\infty(\Omega)$ growth of the coefficients A_n and B_n, $n \ge 0$, may seem quite restrictive. However, these hypotheses have been necessary to relate the $\mathrm{L}^2(\Omega)$ norm of the coefficients $X_0, X_1, X_2, \ldots$ in (3.9), then define the random variables $H_0, H_1, H_2, \ldots$ and finally bound $\|X_n\|_2 \le$

H_n by induction on $n \geq 0$. Without the hypotheses $\|A_n\|_\infty \leq C_r/r^n$ and $\|B_n\|_\infty \leq C_r/r^n$ of Theorem 3.3, this would not have been possible.

Moreover, these $\mathrm{L}^\infty(\Omega)$ hypotheses are equivalent to a growth condition on the moments of the random variables $A_0, A_1, \ldots$ and $B_0, B_1, \ldots$. The key fact is that, for a given random variable Z, we have that $\mathbb{E}[|Z|^n] \leq HR^n$ for certain $H > 0$ and $R > 0$, if and only if $\|Z\|_\infty \leq R$. This key fact is a direct consequence of the following result: if Z is a random variable, then $\lim_{n\to\infty} \|Z\|_n = \|Z\|_\infty$.

Growth hypotheses of the form $\mathbb{E}[|Z|^n] \leq HR^n$, for certain $H > 0$ and $R > 0$, are common in the literature to find stochastic analytic solutions to particular cases of (3.1). See for example Airy's random differential equation in [43] and Hermite's random differential equation in [29]. We have proved in this subsection that controlling the growth of the moments is equivalent to controlling the $\mathrm{L}^\infty(\Omega)$ norm. Hence, Theorem 3.3 will allow us to generalize the results obtained in previous articles, for instance [29, 43]. See Example 3.9 and Example 3.10 for the generalization.

To conclude this subsection, we prove that the hypotheses used in Theorem 3.4 are sharp, in the sense that counterexamples exist if any of them is relaxed. The following two examples of (3.1) with an unbounded input coefficient have no mean square solution $X(t)$. The arguments to prove that these examples have no solution follow the reasoning of [161, Example, pp. 541–542].

Example 3.6 Consider the initial value problem

$$\begin{cases} \ddot{X}(t) + ZX(t) = 0, \ t \in \mathbb{R}, \\ X(t_0 = 0) = Y_0, \\ \dot{X}(t_0 = 0) = 0, \end{cases}$$

where $Z < 0$ is an unbounded random variable (for example, $Z = -U$, where U follows an Exponential, Gamma, Poisson, etc. distribution). Suppose that for any initial condition $Y_0 \in \mathrm{L}^2(\Omega)$ there is a mean square solution $X(t)$. By [161, Th. 3(a)], every mean square solution to a random differential equation problem is a sample-path solution. More specifically, there exists an equivalent stochastic process, product measurable, whose sample paths solve the deterministic counterpart of the problem almost surely. Therefore $X(t)$ is a sample-path solution (we choose the appropriate representative of the equivalence class), with $X(t) = Y_0 \cosh(\sqrt{-Z}\, t)$ for all $t \in \mathbb{R}$, almost surely. Fix $t \neq 0$. Consider the random variable $T = \cosh(\sqrt{-Z}\, t)$. Notice that

$\|T\|_\infty = \infty$. Consider the operator $\Delta : \mathrm{L}^2(\Omega) \to \mathrm{L}^2(\Omega)$, $\Delta(Y) = YT$. This operator is linear and continuous, as a consequence of the closed graph theorem. Hence, there is a constant $C > 0$ such that $\|YT\|_2 \leq C\|Y\|_2$, for all $Y \in \mathrm{L}^2(\Omega)$. In fact, this inequality holds for any random variable Y (since, if $Y \notin \mathrm{L}^2(\Omega)$, then $\|Y\|_2 = \infty$). Let $Y = T^m$. We have $\|T^{m+1}\|_2 \leq C\|T^m\|_2$, which yields $\|T^m\|_2 \leq C^m$. That is, $\|T\|_{2m} \leq C$. Hence, $\|T\|_\infty = \lim_{m\to\infty} \|T\|_{2m} \leq C$, but this is a contradiction. Thus, we conclude that there must exist an initial condition $Y_0 \in \mathrm{L}^2(\Omega)$ such that the stochastic problem has no mean square solution.

The case in which $Z > 0$ is unbounded (let us suppose that Z is Gamma distributed) may be tackled analogously, although with a subtlety. Proceeding again by contradiction, let us suppose that for any initial condition $Y_0 \in \mathrm{L}^2(\Omega)$ there exists a mean square solution $X(t)$. By [161, Th. 3(a)], $X(t) = Y_0 \cos(\sqrt{Z}\, t)$ for all $t \in \mathbb{R}$, almost surely. In contrast with the previous case, now $\cos(\sqrt{Z}\, t)$ is bounded. As $X(t)$ is mean square differentiable, its mean square derivative must be given by $\dot{X}(t) = -Y_0\sqrt{Z}\sin(\sqrt{Z}\, t)$ [56, p. 536]. Fix $t \neq 0$ and let $T = -\sqrt{Z}\sin(\sqrt{Z}\, t)$. Now we do have that $\|T\|_\infty = \infty$, so the previous reasoning based on the closed graph theorem can be applied to deduce that there exists an initial condition $Y_0 \in \mathrm{L}^2(\Omega)$ such that $\dot{X} \notin \mathrm{L}^2(\Omega)$. This is a contradiction.

The general case, in which Z is an unbounded random variable, is easily addressed now (this includes, for instance, the case of Gaussian random variables). If Z is unbounded, then it must be unbounded on the positive or negative axis. Let us suppose it unbounded on the positive axis (the other case is completely analogous). Take $\tilde{\Omega} \subseteq \Omega$ such that $\mathbb{P}[\tilde{\Omega}] > 0$ and $Z(\omega) > 0$ for each $\omega \in \tilde{\Omega}$. Consider the new probability subspace $(\tilde{\Omega}, \mathcal{F}_{\tilde{\Omega}} = \mathcal{F} \cap 2^{\tilde{\Omega}}, \mathbb{P}_{\tilde{\Omega}} = \mathbb{P}|_{\mathcal{F}_{\tilde{\Omega}}})$. We restate the random differential equation problem on this new probability space, where $Z > 0$ is unbounded. The previous case thus applies. Therefore we are done since every mean square solution on Ω must also be a mean square solution on $\tilde{\Omega}$. This analysis terminates the example.

Example 3.7 Let us consider

$$\begin{cases} \ddot{X}(t) + Z\dot{X}(t) = 0, \ t \in \mathbb{R}, \\ X(t_0 = 0) = 0, \\ \dot{X}(t_0 = 0) = Y_1, \end{cases}$$

where Z is any unbounded random variable. Suppose that for any initial condition Y_1, there exists a mean square solution $X(t)$. Let $Y(t) = \dot{X}(t)$,

which satisfies

$$\begin{cases} \dot{Y}(t) + ZY(t) = 0, \ t \in \mathbb{R}, \\ Y(t_0 = 0) = Y_1. \end{cases}$$

By [161, Th. 3(a)], $Y(t) = Y_1 \mathrm{e}^{-Zt}$ for all $t \in \mathbb{R}$, almost surely. Fix $t \neq 0$ and let $T = \mathrm{e}^{-Zt}$. The random variable T is unbounded. Hence, the same reasoning from Example 3.6 based on the closed graph theorem applies again. We conclude that there must exist $Y_1 \in \mathrm{L}^2(\Omega)$ such that $Y(t) \notin \mathrm{L}^2(\Omega)$, which is a contradiction, and we are done with this example.

3.2.4 Comparison with the random differential transform method

We compare our methodology with the random differential transform method proposed in [168]. Given a stochastic process $U(t)$, its random differential transform is defined as

$$\hat{U}(k) = \frac{U^{(k)}(t_0)}{k!}.$$

Its inverse transform is defined as

$$U(t) = \sum_{k=0}^{\infty} \hat{U}(k)(t - t_0)^k.$$

Notice that we are actually considering Taylor series in a random calculus setting. It is formally assumed that the series $\sum_{k=0}^{\infty} \hat{U}(k)(t - t_0)^k$ is mean square convergent on an interval $(t_0 - r, t_0 + r)$, $r > 0$. The computations with the random differential transform method are analyzed in [168, Th. 2.1].

Proposition 3.8 *[168, Th. 2.1] Let $F(t)$ and $G(t)$ be two second-order stochastic processes, with mean square derivatives of k order $F^{(k)}(t)$ and $G^{(k)}(t)$. Then the following results hold:*

(i) If $U(t) = F(t) \pm G(t)$, then $\hat{U}(k) = \hat{F}(k) \pm \hat{G}(k)$.

(ii) If $U(t) = \lambda F(t)$, where λ is a bounded random variable, then $\hat{U}(k) = \lambda \hat{F}(k)$.

(iii) If $U(t) = G^{(m)}(t)$, then $\hat{U}(k) = (k+1) \cdots (k+m) \hat{G}(k+m)$ (here m is a nonnegative integer).

(iv) If $U(t) = F(t)G(t)$, then $\hat{U}(k) = \sum_{n=0}^{k} \hat{F}(n)\hat{G}(k-n)$.

Notice that (iii) and (iv) can be seen as consequences of differentiating random power series (Proposition 3.1) and multiplying random power series (Proposition 3.2), respectively. Thereby, the random transform method is actually the random Fröbenius method. The recursive equations found for $\hat{X}(k)$ are (3.7). Our Theorem 3.3 and Theorem 3.4 give the conditions under which the inverse transform $\sum_{k=0}^{\infty} \hat{X}(k)(t-t_0)^k$ converges.

3.2.5 Statistical information of the solution stochastic process: mean and variance

The expectation and the variance of the stochastic process $X(t) = \sum_{n=0}^{\infty} X_n(t-t_0)^n$ given by (3.6)–(3.7) can be approximated. Indeed, first, one has to obtain X_n as a function of Y_0, Y_1, $A_0, \ldots, A_{n-1}$ and $B_0, \ldots, B_{n-1}$ by recursion via (3.7), for $n = 0, 1, \ldots, N$. After this, we construct a truncation

$$X_N(t) = \sum_{n=0}^{N} X_n(t-t_0)^n \tag{3.14}$$

of the solution stochastic process $X(t)$. Since $X_N(t) \to X(t)$ in $\mathrm{L}^2(\Omega)$ as $N \to \infty$, we have

$$\lim_{N\to\infty} \mathbb{E}[X_N(t)] = \mathbb{E}[X(t)], \quad \lim_{N\to\infty} \mathbb{V}[X_N(t)] = \mathbb{V}[X(t)],$$

see [160, Th. 4.2.1, Th. 4.3.1].

As an example of the manner one can proceed, we show by hand some random coefficients X_n:

$$X_2 = \frac{-1}{2}(A_0Y_1 + B_0Y_0),$$

$$\begin{aligned} X_3 &= \frac{-1}{6}(A_1Y_1 + B_1Y_0 + 2A_0X_2 + B_0Y_1) \\ &= \frac{-1}{6}(A_1Y_1 + B_1Y_0 - A_0^2Y_1 - A_0B_0Y_0 + B_0Y_1), \end{aligned}$$

$$\begin{aligned} X_4 &= \frac{-1}{12}(A_2Y_1 + B_2Y_0 + 2A_1X_2 + B_1X_1 + 3A_0X_3 + B_0X_2) \\ &= \frac{-1}{12}\big(A_2Y_1 + B_2Y_0 - A_1A_0Y_1 - A_1B_0Y_0 + B_1Y_1 - \frac{1}{2}A_0A_1Y_1 - \frac{1}{2}A_0B_1Y_0 \\ &\quad + \frac{1}{2}A_0^3Y_1 + \frac{1}{2}A_0^2B_0Y_0 - \frac{1}{2}A_0B_0Y_1 - \frac{1}{2}B_0A_0Y_1 - \frac{1}{2}B_0^2Y_0\big). \end{aligned}$$

From these computations, we have the truncation (3.14) for $N = 4$, $X_4(t) = Y_0 + Y_1 t + X_2 t^2 + X_3 t^3 + X_4 t^4$. We need to compute $\mathbb{E}[X_4(t)]$ to approximate $\mathbb{E}[X(t)]$. By linearity of the expectation, $\mathbb{E}[X_4(t)] = \mathbb{E}[Y_0] + \mathbb{E}[Y_1]t + \mathbb{E}[X_2]t^2 + \mathbb{E}[X_3]t^3 + \mathbb{E}[X_4]t^4$. Assuming independence of $Y_0, Y_1, A_0, A_1, \ldots, B_0, B_1, \ldots$ and applying a property from [80, p. 93], we are able to compute the expectation of the addends by hand:

$$\mathbb{E}[X_2] = \frac{-1}{2}(\mathbb{E}[A_0]\mathbb{E}[Y_1] + \mathbb{E}[B_0]\mathbb{E}[Y_0]),$$

$$\begin{aligned}\mathbb{E}[X_3] = \frac{-1}{6}(&\mathbb{E}[A_1]\mathbb{E}[Y_1] + \mathbb{E}[B_1]\mathbb{E}[Y_0] - \mathbb{E}[A_0^2]\mathbb{E}[Y_1] \\ &-\mathbb{E}[A_0]\mathbb{E}[B_0]\mathbb{E}[Y_0] + \mathbb{E}[B_0]\mathbb{E}[Y_1]),\end{aligned}$$

$$\begin{aligned}\mathbb{E}[X_4] = \frac{-1}{12}\big(&\mathbb{E}[A_2]\mathbb{E}[Y_1] + \mathbb{E}[B_2]\mathbb{E}[Y_0] - \mathbb{E}[A_1]\mathbb{E}[A_0]\mathbb{E}[Y_1] - \mathbb{E}[A_1]\mathbb{E}[B_0]\mathbb{E}[Y_0] \\ &+\mathbb{E}[B_1]\mathbb{E}[Y_1] - \frac{1}{2}\mathbb{E}[A_0]\mathbb{E}[A_1]\mathbb{E}[Y_1] - \frac{1}{2}\mathbb{E}[A_0]\mathbb{E}[B_1]\mathbb{E}[Y_0] \\ &+\frac{1}{2}\mathbb{E}[A_0^3]\mathbb{E}[Y_1] + \frac{1}{2}\mathbb{E}[A_0^2]\mathbb{E}[B_0]\mathbb{E}[Y_0] - \frac{1}{2}\mathbb{E}[A_0]\mathbb{E}[B_0]\mathbb{E}[Y_1] \\ &-\frac{1}{2}\mathbb{E}[B_0]\mathbb{E}[A_0]\mathbb{E}[Y_1] - \frac{1}{2}\mathbb{E}[B_0^2]\mathbb{E}[Y_0]\big).\end{aligned}$$

For large values of n, we need a computer to manage the big expressions for X_n, and as a consequence $\mathbb{E}[X_N(t)]$ and $\mathbb{V}[X_N(t)]$ for large values of N. We show how to implement the necessary formulas to compute the expectation and variance of the truncated series in the software Mathematica®. The recurrence relation (3.6)–(3.7) is defined as follows:

```
X[n_?NonPositive] := Y0;
X[1] = Y1;
X[n_] := X[n] = -1/(n*(n - 1))*Sum[(m + 1)*A[n - 2 - m]*X[m + 1]
     + B[n - 2 - m]*X[m], {m, 0, n - 2}];
```

Truncation (3.14) is implemented by writing

```
seriesX[t_, t0_, N_] := X[0] + Sum[X[n]*(t - t0)^n, {n, 1, N}];
```

Using the `Expectation` or `NExpectation` function, in which one can set the distributions of `A[n]`, `B[n]`, `Y0` and `Y1`, both the expectation and the variance of (3.14) can be calculated by the computer.

There are other approaches to approximate the expectation of the solution to the random initial value problem (3.1). One of these approaches is the so-called dishonest method [11], [85, p. 149], which assumes that $A(t)$ and $\dot{X}(t)$ are independent and that $B(t)$ and $X(t)$ are independent. Denoting $\mu_X(t) = \mathbb{E}[X(t)]$, the idea is that, since $\mathbb{E}[\ddot{X}(t)] = \frac{\mathrm{d}^2}{\mathrm{d}t^2}(\mu_X(t))$ and $\mathbb{E}[\dot{X}(t)] = \frac{\mathrm{d}}{\mathrm{d}t}(\mu_X(t))$, because of the commutation between the mean square limit and the expectation operator (see [160, Ch. 4]), by the assumed independence we arrive at a deterministic initial value problem to compute $\mu_X(t)$:

$$\begin{cases} \frac{\mathrm{d}^2}{\mathrm{d}t^2}(\mu_X(t)) + \mathbb{E}[A(t)]\frac{\mathrm{d}}{\mathrm{d}t}(\mu_X(t)) + \mathbb{E}[B(t)]\mu_X(t) = 0,\ t \in \mathbb{R}, \\ \mu_X(t_0) = \mathbb{E}[Y_0], \\ \frac{\mathrm{d}}{\mathrm{d}t}(\mu_X(t_0)) = \mathbb{E}[Y_1]. \end{cases} \tag{3.15}$$

In [11] and [85, p. 149] this method is used to handle the problem of computing the expectation of the solution stochastic process of certain random differential equations. In [29, 43] approximations of the expectation of the corresponding solution stochastic process obtained via specific methods have been compared with the ones calculated by the dishonest approach. In our context, the dishonest method will work on cases where $\mathbb{C}\mathrm{ov}[A(t), \dot{X}(t)]$ and $\mathbb{C}\mathrm{ov}[B(t), X(t)]$ is small, but in general, there is no certainty that this may hold. In Example 3.9 and Example 3.10, we approximate $\mathbb{C}\mathrm{ov}[A(t), \dot{X}(t)]$ and $\mathbb{C}\mathrm{ov}[B(t), X(t)]$ in order to understand better the accuracy of the dishonest method. Nevertheless, the approximations via the truncation method previously described allow us to obtain reliable approximations for the expectation, and also for the variance, of the solution process.

Another popular approach, already mentioned in the Introduction chapter, consists in using Monte Carlo simulation. Sample from the distributions of $A(t)$, $B(t)$, Y_0 and Y_1 to obtain, say M realizations, for M large. That is, we have $A(t, \omega_1), \ldots, A(t, \omega_M)$, $B(t, \omega_1), \ldots, B(t, \omega_M)$, $Y_0(\omega_1), \ldots, Y_0(\omega_M)$ and $Y_1(\omega_1), \ldots, Y_1(\omega_M)$, for M outcomes $\omega_1, \ldots, \omega_M \in \Omega$. Then we solve the M deterministic initial value problems

$$\begin{cases} \ddot{X}(t, \omega_i) + A(t, \omega_i)\dot{X}(t, \omega_i) + B(t, \omega_i)X(t, \omega_i) = 0,\ t \in \mathbb{R}, \\ X(t_0, \omega_i) = Y_0(\omega_i), \\ \dot{X}(t_0, \omega_i) = Y_1(\omega_i), \end{cases} \tag{3.16}$$

so that we obtain M realizations of $X(t)$: $X(t, \omega_1), \ldots, X(t, \omega_M)$. The law of large numbers permits approximating $\mathbb{E}[X(t)]$ and $\mathbb{V}[X(t)]$ by computing the sample mean and sample variance of $X(t, \omega_1), \ldots, X(t, \omega_M)$.

Monte Carlo simulation, in contrast to the dishonest method, always gives correct approximations, and as M grows, these approximations are more accurate, although the Monte Carlo method possesses a slow convergence rate, namely, $\mathcal{O}(1/\sqrt{M})$ [172, p. 53]. Thereby, the statistical information computed by means of Monte Carlo simulation must approximate our truncation method.

3.2.6 Obtaining error estimates for the approximation of the solution stochastic process, its mean and its variance

Given an error ϵ, we want to obtain N_ϵ so that $\|X_N(t) - X(t)\|_2 < \epsilon$ for all $N \geq N_\epsilon$. Notice that, in such a case, by Jensen's and Cauchy-Schwarz inequalities we have

$$\begin{aligned} |\mathbb{E}[X_N(t)] - \mathbb{E}[X(t)]| &= |\mathbb{E}[X_N(t) - X(t)]| \\ &\leq \mathbb{E}[|X_N(t) - X(t)|] \leq \|X_N(t) - X(t)\|_2 < \epsilon, \end{aligned}$$

therefore we will be able to estimate the error when approximating the mean $\mathbb{E}[X(t)]$ via $\mathbb{E}[X_N(t)]$.

The method to estimate errors for $\|X_N(t) - X(t)\|_2$ is as follows. We use the notation from the proof of Theorem 3.3. If we denote $\rho = |t - t_0|$ and take $\rho < s < r$, we have

$$\|X_N(t) - X(t)\|_2 = \left\| \sum_{n=N+1}^{\infty} X_n (t - t_0)^n \right\|_2 \leq \sum_{n=N+1}^{\infty} \|X_n\|_2 \rho^n \leq \sum_{n=N+1}^{\infty} H_n \rho^n. \tag{3.17}$$

To bound $H_n s^n$, we base our reasoning on (3.12). Given $M_n = \max_{0 \leq m \leq n} H_m s^m$, we saw in the proof of Theorem 3.3 that $M_n = M$ for sufficiently large n. In fact, for n satisfying

$$\frac{ns}{(n+2)r} + \frac{C_r s}{n+2} + \frac{C_r s^2}{(n+2)(n+1)} < 1, \tag{3.18}$$

it holds $M_n = M$, so M can be computed just by knowing r, C_r, s, $\|Y_0\|_2$ and $\|Y_1\|_2$, because from these values we can see when (3.18) holds and compute H_n via the recursion (3.10) or (3.11), and thereby M. Notice that, if a lot of the random variables A_n and B_n are 0, then H_n will not be a tight bound for $\|X_n\|_2$, in general.

Once we know M, recall from the end of the proof of Theorem 3.3 that $H_n\rho^n \leq M(\rho/s)^n$. We bound from (3.17):

$$\|X_N(t) - X(t)\|_2 \leq M \sum_{n=N+1}^{\infty} \left(\frac{\rho}{s}\right)^n = M\frac{\left(\frac{\rho}{s}\right)^{N+1}}{1-\frac{\rho}{s}}.$$

If we want $\|X_N(t) - X(t)\|_2$ to be smaller than a prefixed error $\epsilon > 0$, we impose

$$M\frac{\left(\frac{\rho}{s}\right)^{N+1}}{1-\frac{\rho}{s}} < \epsilon.$$

This yields

$$N_\epsilon = \left\lceil \frac{\log\left(\frac{\left(1-\frac{\rho}{s}\right)\epsilon}{M}\right)}{\log\left(\frac{\rho}{s}\right)} - 1 \right\rceil \tag{3.19}$$

(here, $\lceil x \rceil$ denotes the least integer that is greater than or equal to x, commonly known as the ceiling of x).

In Example 3.9 and Example 3.12, we will apply these computations to find an N_ϵ for which the approximation of $\mathbb{E}[X(t)]$ via $\mathbb{E}[X_N(t)]$ at the point $t = 1$ and $t = 0.25$, respectively, gives an error smaller than ϵ.

We develop a similar method to estimate the errors in the approximations for the variance. That is to say, given an error $\epsilon > 0$, we want to find N_ϵ such that $|\mathbb{V}[X_N(t)] - \mathbb{V}[X(t)]| < \epsilon$ for all $N \geq N_\epsilon$. We start bounding the difference $|\mathbb{V}[X_N(t)] - \mathbb{V}[X(t)]|$, using triangular, Jensen's and Cauchy-Schwarz inequalities:

$$\begin{aligned}
&|\mathbb{V}[X_N(t)] - \mathbb{V}[X(t)]| = |\mathbb{E}[(X_N(t))^2] - (\mathbb{E}[X_N(t)])^2 - \mathbb{E}[(X(t))^2] + (\mathbb{E}[X(t)])^2| \\
&\leq \mathbb{E}[|(X_N(t))^2 - (X(t))^2|] + |(\mathbb{E}[X_N(t)])^2 - (\mathbb{E}[X(t)])^2| \\
&= \mathbb{E}[|X_N(t) - X(t)||X_N(t) + X(t)|] + |\mathbb{E}[X_N(t)] - \mathbb{E}[X(t)]||\mathbb{E}[X_N(t)] + \mathbb{E}[X(t)]| \\
&\leq \|X_N(t) - X(t)\|_2\|X_N(t) + X(t)\|_2 + |\mathbb{E}[X_N(t)] - \mathbb{E}[X(t)]|(|\mathbb{E}[X_N(t)]| + |\mathbb{E}[X(t)]|) \\
&\leq \|X_N(t) - X(t)\|_2(\|X_N(t)\|_2 + \|X(t)\|_2) + |\mathbb{E}[X_N(t)] - \mathbb{E}[X(t)]|(|\mathbb{E}[X_N(t)]| + |\mathbb{E}[X(t)]|).
\end{aligned}$$

Let $\delta > 0$ that will be determined later on to make the error smaller than ϵ. By the results previously obtained in this subsection, we can choose N_δ such that $\|X_N(t) - X(t)\|_2 < \delta$ for all $N \geq N_\delta$ (just applying (3.19) with $\epsilon = \delta > 0$). Moreover, $|\mathbb{E}[X_N(t)] - \mathbb{E}[X(t)]| \leq \|X_N(t) - X(t)\|_2 < \delta$, by

Jensen's and Cauchy-Schwarz inequalities. Then,

$$\begin{aligned}|\mathbb{V}[X_N(t)] - \mathbb{V}[X(t)]| \leq & \delta(\|X_N(t)\|_2 + \|X(t)\|_2) + \delta(|\mathbb{E}[X_N(t)]| + |\mathbb{E}[X(t)]|) \\ \leq & \delta(\|X_N(t) - X(t) + X(t)\|_2 + \|X(t)\|_2) \\ & + \delta(|\mathbb{E}[X_N(t)] - \mathbb{E}[X_N(t)] + \mathbb{E}[X_N(t)]| + |\mathbb{E}[X(t)]|) \\ \leq & \delta(\|X_N(t) - X(t)\|_2 + \|X(t)\|_2 + \|X(t)\|_2) \\ & + \delta(|\mathbb{E}[X_N(t)] - \mathbb{E}[X_N(t)]| + |\mathbb{E}[X(t)]| + |\mathbb{E}[X(t)]|) \\ \leq & \delta(\delta + 2\|X(t)\|_2) + \delta(\delta + 2|\mathbb{E}[X(t)]|).\end{aligned}$$

To bound $\|X(t)\|_2$, write

$$\begin{aligned}\|X(t)\|_2 = \left\|\sum_{n=0}^{\infty} X_n (t-t_0)^n\right\|_2 \leq & \sum_{n=0}^{\infty} \|X_n\|_2 \rho^n \\ \leq \sum_{n=0}^{\infty} H_n \rho^n \leq M \sum_{n=0}^{\infty} \left(\frac{\rho}{s}\right)^n = & M \frac{1}{1-\frac{\rho}{s}} =: \gamma,\end{aligned}$$

where $\rho = |t - t_0|$ and s is any number satisfying $\rho < s < r$. Before, we saw how to compute M, so we have obtained a computable bound for $\|X(t)\|_2$. On the other hand, to bound $|\mathbb{E}[X(t)]|$ we have two options. One option consists in using Jensen's and Cauchy-Schwarz inequalities to derive $|\mathbb{E}[X(t)]| \leq \|X(t)\|_2 \leq \gamma$, and we are done. The second option, which provides a tighter bound for $|\mathbb{E}[X(t)]|$, consists in using the approximations performed for $\mathbb{E}[X(t)]$ via $\mathbb{E}[X_N(t)]$, and from them deducing an upper bound for $\mathbb{E}[X(t)]$. For any of these two options, we denote the upper bound obtained for $|\mathbb{E}[X(t)]|$ by $\beta > 0$. Thus, for $N \geq N_\delta$,

$$|\mathbb{V}[X_N(t)] - \mathbb{V}[X(t)]| \leq \delta(2\gamma + \delta) + \delta(2\beta + \delta).$$

Now choose δ so that

$$\delta(2\gamma + \delta) + \delta(2\beta + \delta) \leq \epsilon.$$

From here, we take

$$\delta = \frac{-(\gamma + \beta) + \sqrt{(\gamma + \beta)^2 + 2\epsilon}}{2} > 0. \tag{3.20}$$

Thus, to sum up, given a prefixed error $\epsilon > 0$, the steps to be done in order to guarantee the approximations $\mathbb{V}[X_N(t)]$ of the exact variance $\mathbb{V}[X(t)]$ to satisfy $|\mathbb{V}[X_N(t)] - \mathbb{V}[X(t)]| \leq \epsilon$ are the following ones:

1. compute $\gamma = M/(1 - \rho/s)$;

2. compute $\beta > 0$, upper bound of $|\mathbb{E}[X(t)]|$;

3. obtain $\delta > 0$ from (3.20);

4. take N_δ as in the approximation of the mean (expression (3.19), but with δ instead of ϵ).

To put forward these ideas, in Example 3.9 and Example 3.12 we will apply these computations to find an N_ϵ for which, given a *priori* error $\epsilon > 0$, the approximation of $\mathbb{V}[X(t)]$ by means of $\mathbb{V}[X_N(t)]$ at the point $t = 1$ and $t = 0.25$, respectively, gives an error smaller than ϵ.

3.3 Examples

The main goal of this section is to approximate the expectation and variance of the solution process $X(t)$ to particular random initial value problems (3.1). Our tools will be the ones described in Subsection 3.2.5. That is, computing $\mathbb{E}[X_N(t)]$ and $\mathbb{V}[X_N(t)]$ for the truncation (3.14), the dishonest method and Monte Carlo simulation. We will compare the three approaches in order to realize the potentiality of using the truncation method. Our truncation method and Monte Carlo simulation must give similar approximations of the statistical moments (and equal and exact results in the limit).

In addition, we will take some particular problems (3.1) that have been already studied in the literature, as the Airy's and Hermite's random differential equations [43, 29]. As we will see, our findings will generalize the results obtained in those papers (recall Subsection 3.2.3).

Example 3.9 (Airy's random differential equation) Airy-type differential equations appear in a variety of applications to Mathematical Physics, such as the description of the solution to the Schrödinger equation for a particle confined within a triangular potential, in the solution for the one-dimensional motion of a quantum particle affected by a constant force, or in the theory of diffraction of radio waves around the Earth's surface [164]. Airy's random differential equation is given by [43]:

$$\begin{cases} \ddot{X}(t) + AtX(t) = 0, \ t \in \mathbb{R}, \\ X(0) = Y_0, \\ \dot{X}(0) = Y_1, \end{cases} \tag{3.21}$$

where A, Y_0 and Y_1 are random variables.

In [43], the hypothesis used in order to obtain a mean square analytic solution $X(t)$ is $\mathbb{E}[|A|^n] \leq HR^n$, $n \geq n_0$. Notice that this hypothesis is equivalent to $\|A\|_\infty \leq R$, by Subsection 3.2.3.

In the case $\mathbb{E}[|A|^n] \leq HR^n$, $n \geq n_0$ (that is, $\|A\|_\infty \leq R$), we are under the hypotheses of Theorem 3.3. Indeed, in the notation of Theorem 3.3, $A_n = 0$ for all $n \geq 0$, $B_1 = A$ and $B_n = 0$ for every $n \neq 1$. For a fixed and finite $r > 0$ and $t_0 = 0$, we have $\|B_1\|_\infty \leq C_r/r^1$, being $C_r = r\|A\|_\infty$, for instance. Then the stochastic process $X(t) = \sum_{n=0}^{\infty} X_n t^n$, defined as in Theorem 3.3, is a mean square analytic solution to (3.21) in $(-r, r)$. As $r > 0$ is arbitrary, in fact $X(t) = \sum_{n=0}^{\infty} X_n t^n$ is a mean square analytic solution to (3.21) in $\mathbb{R}$.

Let us carry out a practical case. As in [43], consider $A \sim \text{Beta}(2, 3)$ and Y_0, Y_1 independent random variables such that $Y_0 \sim \text{Normal}(1, 1)$ and $Y_1 \sim \text{Normal}(2, 1)$. Table 3.1 and Table 3.2 collect the simulations obtained in [43]. In Table 3.1 we show, for distinct values of t, $\mathbb{E}[X_N(t)]$ for $N = 15$ and $N = 16$, the expectation of the solution stochastic process obtained via the dishonest method and also using Monte Carlo simulation with samples of size $50,000$ and $100,000$. In Table 3.2 we present, for distinct values of t, $\mathbb{V}[X_N(t)]$ for $N = 15$ and $N = 16$ and the corresponding approximations computed via Monte Carlo sampling with $50,000$ and $100,000$ simulations.

We observe that convergence has been achieved for a small N. Compare with Monte Carlo simulation, in which a lot of realizations are required in order to obtain good approximations. Nevertheless, it must be remarked that a small N is needed for the truncation order because Airy's random differential equation is not specially complex. For more complex data processes, as in Example 3.11, Example 3.12 and Example 3.13, a larger order of truncation N may be needed. This may imply a computational expense greater than or similar to Monte Carlo simulation.

On the other hand, it is remarkable how well the dishonest method approximates the correct expectation, although the required independence between $A(t)$, $\dot{X}(t)$ and $B(t)$, $X(t)$ does not hold. The key point is that $\mathbb{C}\text{ov}[A(t), \dot{X}(t)]$ and $\mathbb{C}\text{ov}[B(t), X(t)]$ are small, as Table 3.3 shows, which justifies the accuracy of the dishonest method, especially for small t. Notice that, in this example, $\mathbb{C}\text{ov}[A(t), \dot{X}(t)] = 0$, because $A(t) \equiv 0$ is deterministic. The value of $\mathbb{C}\text{ov}[B(t), X(t)]$ is calculated by considering approximations $X_N(t)$ of $X(t)$ with $N = 16$, since for this order of truncation one gets good approximations of $X(t)$, in other words, approximations can be consider as fairly exact.

t	$\mathbb{E}[X_{15}(t)]$	$\mathbb{E}[X_{16}(t)]$	dishonest	MC $50,000$	MC $100,000$
0.00	1	1	1	0.99701	1.00138
0.25	1.49870	1.49870	1.49870	1.49519	1.49976
0.50	1.98752	1.98752	1.98752	1.98353	1.98829
0.75	2.45108	2.45108	2.45102	2.44667	2.45160
1.00	2.86856	2.86856	2.86818	2.86383	2.86893
1.25	3.21494	3.21494	3.21339	3.21008	3.21534
1.50	3.46310	3.46310	3.45812	3.45831	3.46376
1.75	3.58660	3.58660	3.57340	3.58215	3.58784
2.00	3.56336	3.56335	3.53286	3.55948	3.56552

Table 3.1: Approximation of the expectation of the solution stochastic process. Example 3.9, assuming independent random data.

t	$\mathbb{V}[X_{15}(t)]$	$\mathbb{V}[X_{16}(t)]$	MC $50,000$	MC $100,000$
0.00	1	1	0.99610	0.99530
0.25	1.06035	1.06035	1.05902	1.05642
0.50	1.23142	1.23142	1.23408	1.22793
0.75	1.49261	1.49261	1.50041	1.48944
1.00	1.81392	1.81392	1.82744	1.81127
1.25	2.15870	2.15870	2.17768	2.15721
1.50	2.49379	2.49379	2.51690	2.49462
1.75	2.80560	2.80560	2.83029	2.81030
2.00	3.11530	3.11530	3.13783	3.12559

Table 3.2: Approximation of the variance of the solution stochastic process. Example 3.9, assuming independent random data.

t	$\mathbb{C}\mathrm{ov}[A(t), \dot{X}_{16}(t)]$	$\mathbb{C}\mathrm{ov}[B(t), X_{16}(t)]$
0.00	0	0
0.25	0	−0.0000325384
0.50	0	−0.000622983
0.75	0	−0.00365252
1.00	0	−0.0130099
1.25	0	−0.0349332
1.50	0	−0.0778424
1.75	0	−0.151444
2.00	0	−0.264968

Table 3.3: Approximation of $\mathbb{C}\mathrm{ov}[A(t), \dot{X}(t)]$ and $\mathbb{C}\mathrm{ov}[B(t), X(t)]$ via accurate truncations $\dot{X}_{16}(t)$ and $X_{16}(t)$, respectively. Example 3.9, assuming independent random data.

We perform another example of the Airy's random differential equation (3.21), but this time the input random variables A, Y_0 and Y_1 will not be independent. Indeed, take a random vector (A, Y_0, Y_1) that follows a multivariate Gaussian distribution, with mean vector and covariance matrix

$$\mu = \begin{pmatrix} 0.4 \\ 1 \\ 2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 0.04 & 0.0001 & -0.05 \\ 0.0001 & 1 & 0.5 \\ -0.005 & 0.5 & 1 \end{pmatrix},$$

respectively. In order for the hypotheses of Theorem 3.3 to be satisfied, we need to truncate A (because the Normal distribution is unbounded). Since A follows a Normal distribution with mean $\mu_A = 0.4$ and variance $\sigma_A^2 = 0.04$, the interval $[\mu_A - 3\sigma_A, \mu_A + 3\sigma_A] = [-0.2, 1]$ contains 99.7% of the observations of A. Thus, the multivariate Gaussian distribution will be truncated to $[-0.2, 1] \times \mathbb{R} \times \mathbb{R}$.

In Table 3.4 and Table 3.5, we present the numerical experiments. We use truncation (3.14) with $N = 15$ and $N = 16$, the dishonest method and Monte Carlo simulation.

Once again, convergence has been achieved quite quickly, compared with Monte Carlo simulation. The results obtained are more accurate than via the dishonest method and Monte Carlo simulation. Nonetheless, it is remarkable again the accuracy of the dishonest method, particularly in the time interval $t \in [0, 1]$, although not as good as in the previous case (see Table 3.1). In Table 3.6, we show approximations of the covariances $\mathbb{C}\mathrm{ov}[A(t), \dot{X}(t)]$ and $\mathbb{C}\mathrm{ov}[B(t), X(t)]$. These covariances are small, especially for small t, which explains the good approximation of the expectation via the dishonest method.

t	$\mathbb{E}[X_{15}(t)]$	$\mathbb{E}[X_{16}(t)]$	dishonest	MC $50,000$	MC $100,000$
0.00	1	1	1	1.00287	1.00114
0.25	1.50619	1.50619	1.49870	1.50202	1.50134
0.50	1.98755	1.98755	1.98750	1.99129	1.99168
0.75	2.45120	2.45120	2.45102	2.45529	2.45678
1.00	2.86893	2.86893	2.86818	2.87321	2.87583
1.25	3.23538	3.23538	3.21339	3.22008	3.22386
1.50	3.46485	3.46485	3.45812	3.46878	3.47374
1.75	3.58966	3.58966	3.57340	3.59290	3.59902
2.00	3.56817	3.56817	3.53286	3.57032	3.57755

Table 3.4: Approximation of the expectation of the solution stochastic process. Example 3.9, assuming dependent random data.

t	$\mathbb{V}[X_{15}(t)]$	$\mathbb{V}[X_{16}(t)]$	MC $50,000$	MC $100,000$
0.00	1	1	1.01144	1.00603
0.25	1.31128	1.31128	1.32213	1.31611
0.50	1.75161	1.75161	1.73855	1.73153
0.75	2.22370	2.22370	2.22680	2.21842
1.00	2.72094	2.72094	2.73668	2.72659
1.25	3.06515	3.06515	3.20829	3.19619
1.50	3.57031	3.57031	3.58843	3.57424
1.75	3.83591	3.83590	3.85507	3.83916
2.00	4.02095	4.02100	4.04117	4.02478

Table 3.5: Approximation of the variance of the solution stochastic process. Example 3.9, assuming dependent random data.

t	$\mathbb{C}\mathrm{ov}[A(t), \dot{X}_{16}(t)]$	$\mathbb{C}\mathrm{ov}[B(t), X_{16}(t)]$
0.00	0	0
0.25	0	−0.00106042
0.50	0	−0.00177052
0.75	0	−0.00618508
1.00	0	−0.0173
1.25	0	−0.0509053
1.50	0	−0.0858552
1.75	0	−0.161068
2.00	0	−0.276206

Table 3.6: Approximation of $\mathbb{C}\mathrm{ov}[A(t), \dot{X}(t)]$ and $\mathbb{C}\mathrm{ov}[B(t), X(t)]$ via accurate truncations $\dot{X}_{16}(t)$ and $X_{16}(t)$, respectively. Example 3.9, assuming dependent random data.

As an application of the error estimates studied in Subsection 3.2.6, we estimate for which index N_ϵ the error obtained in the approximation of $\mathbb{E}[X(t)]$ via $\mathbb{E}[X_N(t)]$ is smaller than $\epsilon = 0.00001$. Take $t = 1$ and $r = 2$. Let, for instance, $s = 1.5$. In both cases in this example (assuming independent and dependent random data), we have $C_r = r\|A\|_\infty$. Since $\|A\|_\infty = 1$ (recall that in the two cases considered throughout this example, the realizations $A(\omega)$, $\omega \in \Omega$, of the random variable A lie either in $[0, 1]$ or in $[-0.2, 1]$, thus being less than 1), we take $C_r = r = 2$. From these values, the least n_0 such that (3.18) holds for all $n \geq n_0$, is $n_0 = 7$ (this value is obtained by plotting the left-hand side of (3.18) and looking at the point n_0 from which the graph is less than 1). Then, from (3.12), $M = M_8 = \max_{0\leq m\leq 8} H_m s^m = 2024.49$. Finally, using (3.19), one gets $N_\epsilon = 10$. For $N \geq N_\epsilon = 10$, it holds $|\mathbb{E}[X_N(t)] - \mathbb{E}[X(t)]| < 0.00001$.

Now, given $\epsilon = 0.00001$, we obtain an N_ϵ such that $|\mathbb{V}[X_N(t)] - \mathbb{V}[X(t)]| < \epsilon$ at $t = 1$, for every $N \geq N_\epsilon$. We use the ideas and notation from Subsection 3.2.6. We have $t = 1$, $\rho = 1$, $r = 2$ and $s = 1.5$. We saw that $M = 2024.49$. Then, $\gamma = M/(1 - \rho/s) = 6073.47$. Recall that we could choose β equal to γ or, for a tighter bound, use Table 3.1 and Table 3.4. We see that $|\mathbb{E}[X(t)]| \leq 2.869 =: \beta$. From these values, we obtain $\delta = 8.22638 \cdot 10^{-10}$. Finally, choose N_δ so that $\|X_N(1) - X(1)\|_2 < \delta$. Use formula (3.19) (with δ instead of ϵ) to get $N_\delta = 73$. Thus, for $N \geq 73$, the inequality $|\mathbb{V}[X_N(t)] - \mathbb{V}[X(t)]| < 0.00001$ holds for sure.

Example 3.10 (Hermite's random differential equation) Hermite's random differential equation is defined as follows:

$$\begin{cases} \ddot{X}(t) - 2t\dot{X}(t) + AX(t) = 0, \ t \in \mathbb{R}, \\ X(0) = Y_0, \\ \dot{X}(0) = Y_1, \end{cases} \tag{3.22}$$

where A, Y_0 and Y_1 are random variables.

In [29], the moments of A are controlled as $\mathbb{E}[|A|^n] \leq HR^n$, $n \geq n_0$, to prove the existence of a mean square analytic solution to random initial value problem (3.22). As we saw in Subsection 3.2.3, this hypothesis reduces to $\|A\|_\infty \leq R$.

If $\mathbb{E}[|A|^n] \leq HR^n$, $n \geq n_0$ (that is, $\|A\|_\infty \leq R$), we are under the assumptions of Theorem 3.3. Indeed, in the notation of Theorem 3.3, $A_1 = -2$, $A_n = 0$ for all $n \neq 1$, $B_0 = A$ and $B_n = 0$ for every $n \neq 0$. For a fixed and finite $r > 0$ and $t_0 = 0$, we have $\|A_1\|_\infty \leq C_r/r^1$ and $\|B_0\|_\infty \leq C_r/r^0 = C_r$, being $C_r = \max\{2r, \|A\|_\infty\}$ for example. Then $X(t) = \sum_{n=0}^{\infty} X_n t^n$ defined as in Theorem 3.3 is an analytic solution to (3.22) in $(-r, r)$. Again, as $r > 0$ is arbitrary, $X(t) = \sum_{n=0}^{\infty} X_n t^n$ is a mean square analytic solution stochastic process to random initial value problem (3.22) in $\mathbb{R}$.

As in [29], let $A \sim \text{Normal}(\mu = 5, \sigma^2 = 1)$ and Y_0, Y_1 independent random variables such that $Y_0 \sim \text{Normal}(1, 1)$ and $Y_1 \sim \text{Normal}(2, 1)$. In this case, since the Normal distribution is unbounded, it does not fulfill the hypotheses, therefore we need to truncate it: in [29] it has been truncated to the interval $[\mu - 3\sigma, \mu + 3\sigma] = [2, 8]$, which contains approximately 99.7% of the observations of a Gaussian random variable. Table 3.7 and Table 3.8 simulate the results obtained in [29]. In Table 3.7 we show, for distinct values of t, $\mathbb{E}[X_N(t)]$ for $N = 15$ and $N = 16$, the corresponding approximation obtained via the dishonest method and Monte Carlo simulation with samples of size $50,000$ and $100,000$. In Table 3.8 we present, for distinct values of t, $\mathbb{V}[X_N(t)]$ for $N = 15$ and $N = 16$ and Monte Carlo simulation with samples of size $50,000$ and $100,000$. As it occurred in the previous example, convergence has been achieved for a small N. In Table 3.9, we show approximations of the covariances $\mathbb{C}\text{ov}[A(t), \dot{X}(t)]$ and $\mathbb{C}\text{ov}[B(t), X(t)]$ to understand better the accuracy of the dishonest method.

Example 3.11 (Polynomial data processes) Let us consider more complex data processes in our random differential equation (3.1). The data stochas-

t	$\mathbb{E}[X_{15}(t)]$	$\mathbb{E}[X_{16}(t)]$	dishonest	MC $50,000$	MC $100,000$
0.00	1	1	1	0.99750	0.99919
0.25	1.32907	1.32907	1.32889	1.32619	1.32703
0.50	1.26473	1.26473	1.26175	1.26351	1.26219
0.75	0.74510	0.74510	0.72906	0.74737	0.74316
1.00	−0.27157	−0.27157	−0.32467	−0.26484	−0.27144
1.25	−1.80636	−1.80635	−1.93991	−1.79597	−1.80237
1.50	−3.85882	−3.85868	−4.13681	−3.84872	−3.84906
1.75	−6.40911	−6.40754	−6.90081	−6.40873	−6.39186
2.00	−9.43553	−9.42222	−10.1448	−9.46498	−9.41066

Table 3.7: Approximation of the expectation of the solution stochastic process. Example 3.10.

t	$\mathbb{V}[X_{15}(t)]$	$\mathbb{V}[X_{16}(t)]$	MC $50,000$	MC $100,000$
0.00	1	1	0.98671	1.00327
0.25	0.77433	0.77433	0.76716	0.77821
0.50	0.37752	0.37752	0.37822	0.37992
0.75	0.54181	0.54181	0.53554	0.54357
1.00	2.10396	2.10396	2.06444	2.10993
1.25	5.48674	5.48670	5.40047	5.50378
1.50	10.4476	10.4467	10.3456	10.4828
1.75	18.0186	18.0108	17.9539	18.0963
2.00	43.5731	43.6462	43.5773	44.1340

Table 3.8: Approximation of the variance of the solution stochastic process. Example 3.10.

t	$\mathbb{C}\mathrm{ov}[A(t), \dot{X}_{16}(t)]$	$\mathbb{C}\mathrm{ov}[B(t), X_{16}(t)]$
0.00	0	0
0.25	0	−0.0345004
0.50	0	−0.145742
0.75	0	−0.322733
1.00	0	−0.515142
1.25	0	−0.612961
1.50	0	−0.403919
1.75	0	0.522644
2.00	0	3.0078

Table 3.9: Approximation of $\mathbb{C}\mathrm{ov}[A(t), \dot{X}(t)]$ and $\mathbb{C}\mathrm{ov}[B(t), X(t)]$ via accurate truncations $\dot{X}_{16}(t)$ and $X_{16}(t)$, respectively. Example 3.10.

tic processes will be random polynomials. For example,

$$\begin{cases} \ddot{X}(t) + (A_0 + A_1 t)\dot{X}(t) + (B_0 + B_1 t)X(t) = 0, \ t \in \mathbb{R}, \\ X(0) = Y_0, \\ \dot{X}(0) = Y_1, \end{cases} \tag{3.23}$$

where $A_0 = 4$, $A_1 \sim \mathrm{Uniform}(0, 1)$, $B_0 \sim \mathrm{Gamma}(2, 2)$, $B_1 \sim \mathrm{Bernoulli}(0.35)$, $Y_0 = -1$ and $Y_1 \sim \mathrm{Binomial}(2, 0.29)$ are assumed to be independent. In order for the hypotheses of Theorem 3.3 to be satisfied, the Gamma distribution will be truncated. For the Gamma distribution with shape and rate 2, it can straightforwardly be checked that the interval $[0, 4]$ contains approximately 99.7% of the observations.

By Theorem 3.3, the mean square solution of (3.23) can be written as a random power series $X(t) = \sum_{n=0}^{\infty} X_n t^n$ that is mean square convergent for all $t \in \mathbb{R}$.

In Table 3.10 and Table 3.11, the numerical experiments for the expectation and variance are presented.

Example 3.12 (Infinite series data processes) In this example, the data processes in the random differential equation (3.1) are non-polynomial analytic stochastic process:

$$\begin{cases} \ddot{X}(t) + A(t)\dot{X}(t) + B(t)X(t) = 0, \ t \in \mathbb{R}, \\ X(0) = Y_0, \\ \dot{X}(0) = Y_1, \end{cases} \tag{3.24}$$

t	$\mathbb{E}[X_{19}(t)]$	$\mathbb{E}[X_{20}(t)]$	dishonest	MC $50,000$	MC $100,000$
0.00	-1	-1	-1	-1	-1
0.25	-0.886467	-0.886467	-0.886418	-0.886789	-0.886432
0.50	-0.809269	-0.809269	-0.808743	-0.809370	-0.809219
0.75	-0.747589	-0.747589	-0.745742	-0.747321	-0.747526
1.00	-0.693453	-0.693453	-0.689284	-0.692816	-0.693375
1.25	-0.643943	-0.643944	-0.636462	-0.642985	-0.643845
1.50	-0.598053	-0.598082	-0.586360	-0.596854	-0.597952

Table 3.10: Approximation of the expectation of the solution stochastic process. Example 3.11.

t	$\mathbb{V}[X_{15}(t)]$	$\mathbb{V}[X_{16}(t)]$	MC $50,000$	MC $100,000$
0.00	0	0	0	0
0.25	0.0102077	0.0102074	0.0101172	0.0102664
0.50	0.0190996	0.0190999	0.0189214	0.0192053
0.75	0.0237400	0.0237403	0.0235191	0.0238499
1.00	0.0268721	0.0268711	0.0266311	0.0269620
1.25	0.0297852	0.0297465	0.0295049	0.0298201
1.50	0.0333309	0.0325867	0.0325021	0.0328009

Table 3.11: Approximation of the variance of the solution stochastic process. Example 3.11.

where $A_n \sim \text{Beta}(11, 15)$, for $n \geq 0$, $B_n = 1/n^2$, for $n \geq 1$, and $Y_0 \sim \text{Poisson}(2)$ and $Y_1 \sim \text{Uniform}(0, 1)$ are assumed to be independent. We have $\mathbb{E}[A_n] = 11/26$ and $\mathbb{V}[A_n] = 55/6084$, therefore $\|A_n\|_2 = \sqrt{\mathbb{V}[A_n] + \mathbb{E}[A_n]^2} = 0.12908$, $n \geq 0$. Then

$$\sum_{n=0}^{\infty} \|A_n\|_2 t^n = 0.12908 \sum_{n=0}^{\infty} t^n,$$

which is convergent for $t \in (-1, 1)$. On the other hand,

$$\sum_{n=0}^{\infty} \|B_n\|_2 t^n = \sum_{n=1}^{\infty} \frac{1}{n^2} t^n,$$

which is convergent for $t \in (-1, 1)$ as well. Therefore, the maximum r we can take so that the random differential equation (3.24) makes sense is $r = 1$.

Since $|A_n(\omega)| \leq 1$ for all $\omega \in \Omega$, $|B_n| \leq 1$ and $r = 1$, we can take $C_r = 1$ in Theorem 3.3 and the hypotheses hold. By Theorem 3.3, the mean square solution of (3.24), $X(t)$, is defined and is mean square analytic on $(-1, 1)$.

In Table 3.12 and Table 3.13 we present the numerical experiments. To apply the dishonest method, we need the following two computations:

$$\mathbb{E}[A(t)] = \frac{11}{26} \sum_{n=0}^{\infty} t^n = \frac{11}{26(1-t)}, \quad t \in (-1, 1),$$

and

$$\mathbb{E}[B(t)] = \sum_{n=1}^{\infty} \frac{t^n}{n^2}, \quad t \in (-1, 1).$$

To apply Monte Carlo simulation, we need realizations of the stochastic process $A(t)$, that is, realizations of the random variables $A_0, A_1, \ldots$. As we cannot obtain infinite realizations of a Beta distribution in the computer, we will approximate $A(t, \omega) \approx \sum_{n=0}^{100} A_n(t, \omega) t^n$, so from realizations of $A_0, \ldots, A_{100}$, we will obtain an approximation of a realization of $A(t)$.

By contrast, our approximations using truncation $X_N(t)$, $t \in (-1, 1)$, do not require realizations of the infinite data stochastic process $A(t)$.

As it is observed in Table 3.12 and Table 3.13, the convergence has been practically achieved for $N = 17$.

As an application of the error estimates analyzed in Subsection 3.2.6, we estimate for which index N_ϵ the error obtained in the approximation of $\mathbb{E}[X(t)]$

t	$\mathbb{E}[X_{16}(t)]$	$\mathbb{E}[X_{17}(t)]$	dishonest	MC $50,000$	MC $100,000$
0.00	2	2	2	2.01406	1.99858
0.25	2.11267	2.11267	2.11266	2.12676	2.11130
0.50	2.17662	2.17662	2.17654	2.19050	2.17531
0.75	2.15693	2.15693	2.15675	2.17021	2.15573

Table 3.12: Approximation of the expectation of the solution stochastic process. Example 3.12, assuming independent initial conditions.

t	$\mathbb{V}[X_{17}(t)]$	$\mathbb{V}[X_{18}(t)]$	MC $50,000$	MC $100,000$
0.00	2	2	2.00274	2.00822
0.25	1.99421	1.99421	1.99725	2.00356
0.50	1.93408	1.93408	1.93725	1.94407
0.75	1.76917	1.76919	1.77222	1.77899

Table 3.13: Approximation of the variance of the solution stochastic process. Example 3.12, assuming independent initial conditions.

via $\mathbb{E}[X_N(t)]$ is smaller than $\epsilon = 0.00001$. Take $t = 0.25$ and, for instance, $s = 0.5$. We have $r = 1$ and $C_r = 1$, and from these values we obtain the least n_0 such that (3.18) holds for all $n \geq n_0$, by trial and error. We obtain $n_0 = 0$. From (3.12), $M = M_1 = \max\{H_0, H_1 s\} = \max\{\|Y_0\|_2, \|Y_1\|_2 \cdot 0.5\} = \max\{\sqrt{6}, 1/(2\sqrt{3})\} = \sqrt{6}$, whence $N_\epsilon = 18$ by using (3.19).

Now we bound the error made when approximating the variance. Given an error $\epsilon = 0.00001$, we obtain an N_ϵ such that $|\mathbb{V}[X_N(t)] - \mathbb{V}[X(t)]| < \epsilon$ at the point $t = 0.25$, for all $N \geq N_\epsilon$. We use the ideas and notation from Subsection 3.2.6. We have $t = 0.25$, $\rho = 0.25$, $r = 1$ and $s = 0.5$. We computed $M = \sqrt{6}$, whence $\gamma = M/(1 - \rho/s) = 4.89898$. Recall that we could choose β equal to γ or, for a tighter bound, use Table 3.12. In Table 3.12, we see that $|\mathbb{E}[X(0.25)]| \leq 2.113 =: \beta$. From these values, $\delta = 7.13065 \cdot 10^{-7}$. To end up, pick N_δ such that $\|X_N(0.25) - X(0.25)\|_2 < \delta$. Using expression (3.19) (with δ instead of ϵ) we get $N_\delta = 22$. Thereby, $|\mathbb{V}[X_N(0.25)] - \mathbb{V}[X(0.25)]| < 0.00001$ for $N \geq 22$.

We perform another example for the random initial value problem (3.24), again with $A_n \sim \text{Beta}(11, 15)$, for $n \geq 0$, $B_n = 1/n^2$, for $n \geq 1$, but now the random vector (Y_0, Y_1) follows a Multinomial distribution with 3 repetitions and probabilities 0.29 and 0.15. The random variables/vectors $A_0, A_1, \ldots$ and

(Y_0, Y_1) are independent, but, obviously, Y_0 and Y_1 are not independent. Again, the solution stochastic process $X(t)$ is defined on $(-1, 1)$.

In Table 3.14 and Table 3.15, we show the numerical experiments. Convergence has been practically achieved.

t	$\mathbb{E}[X_{15}(t)]$	$\mathbb{E}[X_{16}(t)]$	dishonest	MC $50,000$	MC $100,000$
0.00	0.87	0.87	0.87	0.866780	0.869210
0.25	0.973828	0.973828	0.973819	0.970205	0.973499
0.50	1.04817	1.04817	1.0481	1.04427	1.04823
0.75	1.07358	1.07358	1.0734	1.06961	1.07394

Table 3.14: Approximation of the expectation of the solution stochastic process. Example 3.12, assuming dependent initial conditions.

t	$\mathbb{V}[X_{14}(t)]$	$\mathbb{V}[X_{15}(t)]$	MC $50,000$	MC $100,000$
0.00	0.617700	0.617700	0.616645	0.621350
0.25	0.574373	0.574373	0.574355	0.577586
0.50	0.553754	0.553754	0.554522	0.556844
0.75	0.525683	0.525704	0.526991	0.528727

Table 3.15: Approximation of the variance of the solution stochastic process. Example 3.12, assuming dependent initial conditions.

Example 3.13 (Truncation method and Monte Carlo for modeling) In order to see a real application of our theoretical development, let us fit data that describe the fish weight growth over the time via a random second-order linear differential equation. In Figure 3.1, we show the fish weight in lbs (vertical axis) per year (horizontal axis). The fish weight datum at the i-th year will be denoted by w_i, for $1 \leq i \leq 33$.

These data have been previously used in [39], where a randomized Bernoulli differential equation has been used, taking as reference Bertalanffy model [146, p. 331].

Let W be the stochastic process that models the fish weight. The random variable $W(t)$ models the fish weight at year t, $1 \leq t \leq 33$. Since $W(t)$ is a positive random variable, we will work with $X(t) = \log(W(t))$, instead. The observed data become $\log(w_1), \ldots, \log(w_{33})$. We use the random initial value

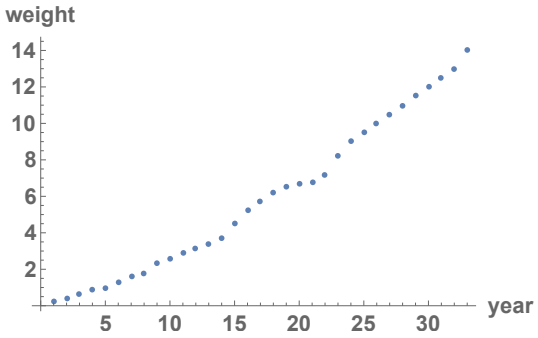

Figure 3.1: Data on fish weights. In the horizontal axis, we represent the years, from 1 to 33. In the vertical axis, we represent the weights in lbs.

problem

$$\begin{cases} \ddot{X}(t) + A_0\dot{X}(t) + (B_0 + B_1 t)X(t) = 0, \ t \in \mathbb{R}, \\ X(0) = Y_0, \\ \dot{X}(0) = Y_1 \end{cases} \tag{3.25}$$

to model the logarithm of the fish weight growth. The stochastic processes $A(t) = A_0$ and $B(t) = B_0 + B_1 t$ have been chosen in such a way by numerical fit trials and computational viability. Notice that (3.25) is a generalization of Airy's random differential equation from Example 3.9.

Using the data drawn in Figure 3.1, we would like to find the best random variables A_0, B_0, B_1, Y_0 and Y_1 so that $W(t)$ fits appropriately the uncertainty associated to the fish weight growth. Since we do not have an explicit solution process X, we will use truncation (3.14), $X_N(t)$, to approximate it in the $\mathrm{L}^2(\Omega)$ sense. Using the truncation with a high N, we will be able to give a suitable distribution for $(A_0, B_0, B_1, Y_0, Y_1)$.

There are two statistical approaches to deal with this problem: the frequentist and the Bayesian techniques. Reference [112] provides an introduction to Bayesian statistics. We do not carry out a Bayesian approach, because $X_N(t)$ has a very large expression, which makes the use of Bayesian estimation impracticable in the computer. Thereby, we use the ideas of the so-called inverse frequentist technique for parameter estimation exhibited in [39] and [157, Ch. 7]. In order not to depart from our reasoning, we will explain our concrete frequentist approach in Remark 3.14 at the end of this example.

Without entering into the theoretical details that will be explained in Remark 3.14, we specify the steps to solve our modeling problem. Computational viability makes us to choose $N = 24$ as the order of truncation. We give the random vector $(A_0, B_0, B_1, Y_0, Y_1)$ a sixth dimensional multinormal random distribution (in the end, it will be truncated so that the hypotheses of Theorem 3.3 fulfill). The mean vector μ is determined as the solution of the deterministic minimization problem

$$\min_{a_0,\, b_0,\, b_1,\, y_0,\, y_1 \in \mathbb{R}} \sum_{i=1}^{33} \left(\log(w_i) - X_{24}(t_i|a_0, b_0, b_1, y_0, y_1)\right)^2,$$

where $X_{24}(t|a_0, b_0, b_1, y_0, y_1)$ corresponds to the value of $X_{24}(t)$ substituting A_0, B_0, B_1, Y_0 and Y_1 by the real numbers a_0, b_0, b_1, y_0 and y_1. This minimization problem can be solved with the built-in function `FindFit` with the option `Method -> NMinimize` in Mathematica®. We obtain

$$\mu = \begin{pmatrix} 0.169695 \\ -0.0123653 \\ 0.000347771 \\ -2.09309 \\ 0.672599 \end{pmatrix}.$$

The covariance matrix Σ is estimated with $\sigma^2 (J^T J)^{-1}$, where

$$\sigma^2 = \frac{\sum_{i=1}^{33} \left(\log(w_i) - X_{24}(t_i|\mu)\right)^2}{33 - 5} = 0.00369358,$$

and J is the Jacobian matrix of $X_{24}(t)$ with respect to $(A_0, B_0, B_1, Y_0, Y_1)$ evaluated at μ. We obtain

$$\Sigma = \begin{pmatrix} 0.000109461 & -0.000010458 & 1.44986 \cdot 10^{-7} & -0.000645456 & 0.000435313 \\ -0.000010458 & 2.47732 \cdot 10^{-6} & -7.94312 \cdot 10^{-8} & 0.0000461867 & -0.0000398354 \\ 1.44986 \cdot 10^{-7} & -7.94312 \cdot 10^{-8} & 3.23082 \cdot 10^{-9} & -3.16049 \cdot 10^{-7} & 5.53045 \cdot 10^{-7} \\ -0.000645456 & 0.0000461867 & -3.16049 \cdot 10^{-7} & 0.00624016 & -0.00312264 \\ 0.000435313 & -0.0000398354 & 5.53045 \cdot 10^{-7} & -0.00312264 & 0.00186506 \end{pmatrix}.$$

Once we know the estimated distribution of $(A_0, B_0, B_1, Y_0, Y_1)$, we know the distribution of $X_{24}(t)$, at least theoretically. The computational complexity of $X_{24}(t)$ is very big, so computing its exact expectation or even good approximations of it is nearly impossible. Due to the complexity of both the truncation expression and the distribution of $(A_0, B_0, B_1, Y_0, Y_1)$, it is better to perform Monte Carlo simulation directly on (3.25) via simulations of $(A_0, B_0, B_1, Y_0, Y_1)$, which follows a (truncated) multivariate Gaussian distribution (μ, Σ).

By means of Monte Carlo simulation with $100,000$ iterates, we obtain samples of $X(i)$, $i = 1, \ldots, 33$. Applying exponential, we have samples of $W(i)$, $i = 1, \ldots, 33$. Hence, approximations of both $\mathbb{E}[W(i)]$ and $\mathbb{V}[W(i)]$ can be calculated. A confidence interval can be computed in two ways: either considering $[\mathbb{E}[W(t)] \pm 2\sqrt{\mathbb{V}[W(t)]}]$ (this is based on how confidence intervals are constructed in a Gaussian setting) or obtaining an accurate approximation using the quartiles of the sample produced by Monte Carlo. In Figure 3.2 and Figure 3.3, the results are shown. As it is observed in both plots, the mean approximates well the real data. However, the confidence interval grows as we move away from 0. Intuitively, this may hold because a truncated random power series centered at t_0 works better near t_0. This phenomenon may be resolved by making the order of truncation N larger and larger (if the computer permits it).

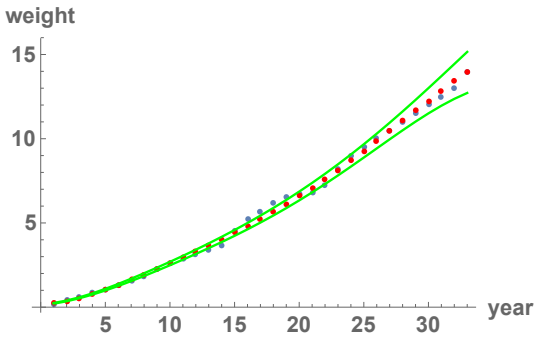

Figure 3.2: Fit of the fish weight data. The blue points represent the real weights, the red points represent the estimated weights (the mean) and the green lines cover a 95% confidence interval constructed with the Gaussian rule [mean $\pm\, 2 \cdot$ standard deviation].

Remark 3.14 (Frequentist technique used in Example 3.13) Let X be a random vector of size n to be modeled (in our case, $(\log W(1), \ldots, \log W(33))$). We set a model of the form $X = f(V)$, where V is a random vector with p components (in our case $(A_0, B_0, B_1, Y_0, Y_1)$) that follows a multinormal distribution with parameters (μ, Σ), and $f : \mathbb{R}^p \to \mathbb{R}^n$ is a function, maybe nonlinear (in our case, $n = 33$ and $f_i(a_0, b_0, b_1, y_0, y_1) = X_{24}(i|a_0, b_0, b_1, y_0, y_1)$, $i = 1, \ldots, 33$, where $X_{24}(t|a_0, b_0, b_1, y_0, y_1)$ is the truncation that approximates the solution of the random differential equation). Let $x = X(\omega)$ be a vector realization of X (in our example, the real data $x = (\log(w_1), \ldots, \log(w_{33}))$). From x, we want to estimate the best μ and Σ so that the model $X = f(V)$

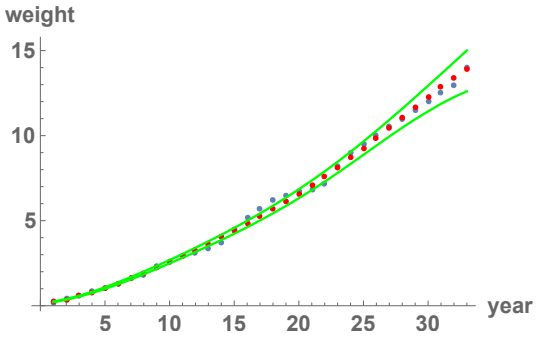

Figure 3.3: Fit of the fish weight data. The blue points represent the real weights, the red points represent the estimated weights (the mean) and the green lines cover a 95% confidence interval constructed by taking the quartiles in the Monte Carlo sampling.

can be considered correct. Let $\hat{v} \in \mathbb{R}^p$ be the minimizer of

$$\min_{v \in \mathbb{R}^p} \sum_{i=1}^{n} (x_i - f_i(v))^2.$$

Using Taylor's expansion,

$$X \approx f(\hat{v}) + Jf(\hat{v})(V - \hat{v}),$$

where J stands for the Jacobian. Then

$$Z := X - f(\hat{v}) + Jf(\hat{v})\hat{v} \approx \underbrace{Jf(\hat{v})}_{J} V.$$

We derive that $Z \approx J\mu + E$, where E follows a multivariate normal distribution with parameters $(0, J\Sigma J^T)$ (here T stands for the transpose matrix operator), i.e., $E \sim \mathrm{MN}(0, J\Sigma J^T)$. Write $J\Sigma J^T = P^T D P$, where P and D are an orthogonal and a diagonal matrix, respectively. Multiplying by P, we have $\bar{Z} \approx \bar{J}\mu + \bar{E}$, where $\bar{Z} = PZ$, $\bar{J} = PJ$ and $\bar{E} = PE \sim \mathrm{MN}(0, D)$. Therefore, $\bar{Z} \approx \bar{J}\mu + \bar{E}$ is a classical linear model (see [134, Ch. 7], [157, Ch. 7]) with normal and independent errors. In a linear model, one should assume homoscedasticity so that the estimation of the parameters makes sense. Thus, we impose $D = \sigma^2 I_n$, where σ^2 is the variance of the errors in the linear model. As a consequence, $\bar{Z} \approx \bar{J}\mu + \bar{E}$ is a classical linear model with homoscedasticity, and the estimations follow from general theory: $\hat{\mu}$ is the minimizer of

$$\min_{\mu \in \mathbb{R}^p} \|\bar{z} - \bar{J}\mu\|_2^2 = \min_{\mu \in \mathbb{R}^p} \|Pz - PJ\mu\|_2^2 = \min_{\mu \in \mathbb{R}^p} \|z - J\mu\|_2^2,$$

where $z = x - f(\hat{v}) + J\hat{v}$ is the vector realization of Z, $\bar{z} = Pz$ is the vector realization of $\bar{Z}$ and $\|\cdot\|_2$ is the euclidean norm. Now,

$$\|z - J\mu\|_2^2 = \|x - (f(\hat{v}) + J(\mu - \hat{v}))\|_2^2 \approx \|x - f(\mu)\|_2^2,$$

so we can take $\hat{\mu} = \hat{v}$, which justifies our choice for the mean in the example. On the other hand, by the general theory of linear models,

$$\hat{\sigma}^2 = \frac{\|\bar{z} - \bar{J}\hat{\mu}\|_2^2}{n-p} = \frac{\|Pz - PJ\hat{\mu}\|_2^2}{n-p} = \frac{\|z - J\hat{\mu}\|_2^2}{n-p} \approx \frac{\|x - f(\hat{\mu})\|_2^2}{n-p}.$$

Finally, from $J\Sigma J^T = P^T DP = \sigma^2 P^T P = \sigma^2 I_n$, we derive $\Sigma = \sigma^2 (J^T J)^{-1}$, by multiplying by $(J^T J)^{-1} J^T$ to the left and by $J(J^T J)^{-1}$ to the right at both sides of the equality (we assume $\operatorname{rank}(J) = p$ so that $(J^T J)^{-1}$ exists). Thus, we choose the estimator $\hat{\Sigma} = \hat{\sigma}^2 (J^T J)^{-1}$.

3.4 Conclusions

In this chapter we have determined analytic stochastic processes that are solutions to the random non-autonomous second-order linear differential equation in the mean square sense. After reviewing the $\mathrm{L}^p(\Omega)$ random calculus and results concerning random power series (differentiation of a random power series in the $\mathrm{L}^p(\Omega)$ sense and Merten's Theorem for random series in the mean square sense), we stated the main theorem of the chapter, Theorem 3.3 and 3.4. These results give assumptions on the coefficient stochastic processes and on the random initial conditions of a random non-autonomous second-order linear differential equation, so that there exists an analytic stochastic process that is a solution in the mean square sense. The mean square approach permitted approximating the main statistical information of the solution, the expectation and the variance. These approximations for the expectation and the variance were compared with other methods previously used in the literature: the dishonest method and Monte Carlo simulation.

The numerical examples presented illustrate the potentiality of our results. The examples show that our findings allow for much more complex random non-autonomous second-order linear differential equations than those from the extant literature. The ideas of this chapter permit dealing with any random non-autonomous second-order linear differential equation in a general form. The statistical information of the stochastic solution can be computed up to any degree of accuracy.

Moreover, our truncation method provides a methodology to estimate the parameters of the multivariate normally distributed explanatory random vector,

in the modeling of real data via random non-autonomous second-order linear differential equations. This procedure together with the Monte Carlo simulation give fitting approximations of the real data.

We mention that a future line of research could consist in extending our ideas to a random non-autonomous linear differential equation of order higher than two. One would need all input stochastic coefficients to be random power series in $L^{\infty}(\Omega)$, in analogy to the hypotheses of Theorem 3.4.

We could also try to apply the random Fröbenius method to the random Riccati differential equation with analytic input coefficients. In [168, Section 3], the authors applied the random differential transform method (which is equivalent to a formal random Fröbenius method, as seen in this chapter) to a particular case of the random Riccati differential equation with a random autonomous coefficient term. It would be interesting to apply the random Fröbenius method in the situation in which all input coefficients are analytic stochastic processes, by proving theoretical results and performing numerical experiments.

Acknowledgements

This work has been supported by the Spanish Ministerio de Economía y Competitividad grant MTM2017–89664–P. I acknowledge the doctorate scholarship granted by Programa de Ayudas de Investigación y Desarrollo (PAID), Universitat Politècnica de València.

The main results of this chapter have been published in [24, 25]. This chapter is an international collaboration with the researcher Laura Villafuerte Altúzar (Department of Mathematics, University of Texas at Austin, USA).

Chapter 4

Beyond the hypothesis of boundedness for the random coefficient of Airy, Hermite and Laguerre differential equations with uncertainties

In this chapter, we study the full randomized versions of Airy, Hermite and Laguerre differential equations, which depend on a random variable appearing as an equation coefficient as well as two random initial conditions. In previous contributions, the mean square stochastic solutions to the aforementioned random differential equations were constructed via the Fröbenius method, under the assumption of exponential growth of the absolute moments of the equation coefficient, which is equivalent to its essential boundedness. In this chapter we aim at relaxing the boundedness hypothesis to allow more general probability distributions for the equation coefficient. We prove that the equations are solvable in the mean square sense when the equation coefficient has finite moment-generating function in a neighborhood of the origin. A thorough discussion of the new hypotheses is included.

4.1 Introduction

In the recent years, important efforts have been made in the analysis of random second-order linear differential equations, in order to extend their deterministic counterpart. The main goal has been to construct the rigorous mean square stochastic solution, thus obtaining approximations for its mean and its variance. Important random differential equations from Mathematical Physics that have been solved are Airy's equation,

$$\ddot{X}(t) + AtX(t) = 0,\ t \in \mathbb{R}, \quad X(0) = Y_0, \quad \dot{X}(0) = Y_1, \tag{4.1}$$

Hermite's equation,

$$\ddot{X}(t) - 2t\dot{X}(t) + AX(t) = 0,\ t \in \mathbb{R}, \quad X(0) = Y_0, \quad \dot{X}(0) = Y_1, \tag{4.2}$$

and Laguerre's equation,

$$t\ddot{X}(t) + (1-t)\dot{X}(t) + AX(t) = 0,\ t \in \mathbb{R}, \quad X(0) = Y_0. \tag{4.3}$$

It is assumed that the equation coefficient A and the initial conditions Y_0 and Y_1 are random variables on an underlying complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For the sake of generality, no statistical independence is assumed between A, Y_0 and Y_1. The term $X(t)$ is a stochastic process, which represents the mean square solution.

In [43, 29, 44], the Fröbenius method was utilized to find a random power series solution $X(t) = \sum_{n=0}^{\infty} X_n t^n$, i.e., a mean square analytic solution, to Airy's, Hermite's and Laguerre's equation, respectively. The coefficients X_n are second-order random variables satisfying $\sum_{n=0}^{\infty} \|X_n\|_2 |t|^n < \infty$, for $t \in \mathbb{R}$. A common hypothesis in those works was the exponential growth of the absolute moments of A: $\mathbb{E}[|A|^n] \leq \eta \mathcal{H}^n$ for $n \geq 1$, for certain constants $\eta, \mathcal{H} > 0$. This assumption is equivalent to the boundedness of A, $\|A\|_\infty < \infty$, as shown in Chapter 2. Under boundedness of the equation coefficients, in Chapter 3 we studied general random second-order linear differential equations, $\ddot{X}(t) + C(t)\dot{X}(t) + D(t)X(t) = 0$, by constructing the mean square convergent power series solution via the Fröbenius method.

The boundedness assumption for A is quite general. Indeed, many standard probability distributions have bounded support: Beta, Triangular, Uniform, Binomial, etc. Moreover, the theorem of existence and uniqueness of mean square solution to general random initial value problems, which is an extension of the classical Picard's theorem with Lipschitz assumption, requires boundedness of the equation coefficients [160, Ch. 5]. By truncating unbounded supports one can use truncated Normal, Gamma and Poisson distributions,

for instance. However, it would be interesting to prove the existence of mean square solution under a more general assumption than boundedness. We take as main reference here the works [28, 30, 27]. Reference [28] studies random first-order linear differential equations, [30] is devoted to a specific random second-order linear differential equation to introduce some random trigonometric functions, and [27] is a doctoral dissertation on the application of the Fröbenius method to solve random differential equations. Following these works, we will assume that Y_0 and Y_1 are fourth-order random variables and that

$$\|A^n\|_4 \leq \eta \mathcal{H}^{n-1}(n-1)!^p \tag{4.4}$$

for $n \geq n_0$, for constants $n_0, \eta, \mathcal{H}, p > 0$. This includes the case of A being bounded, but also the Normal ($p = 1/2$) and the Gamma ($p = 1$) distributions. The assumption (4.4) is a straightforward consequence of polynomial growth for the ratio of moments: $\|A^{n+1}\|_4/\|A^n\|_4 = \mathcal{O}(n^p)$, where the constant corresponding to $\mathcal{O}$ is $\mathcal{H}$, and $\eta = \|A\|_4$.

For Airy's, Hermite's and Laguerre's equation, we will prove that the formal power series solution $X(t) = \sum_{n=0}^{\infty} X_n t^n$ constructed in [43, 29, 44] satisfies $\sum_{n=0}^{\infty} \|X_n\|_2 |t|^n < \infty$, for $t \in \mathbb{R}$. This is enough since the random power series can be differentiated termwise in the mean square sense (Theorem 3.1 in Chapter 3). Our reasoning will be more concise and optimized than in such contributions.

In the last part of the chapter, we will include a thorough discussion about the new hypotheses that permit extending the results published in [43, 29, 44]. We will show the equivalence between condition (4.4) for $p \leq 1$ and $\phi_A(t) < \infty$ in a neighborhood of 0, where $\phi_A(t) = \mathbb{E}[\mathrm{e}^{tA}]$ is the moment-generating function of A. This includes a lot of important probability distributions for A.

4.2 Random Airy differential equation

The formal solution to (4.1) is given by

$$X(t) = Y_0 X_1(t) + Y_1 X_2(t),$$

$$X_1(t) = 1 + \sum_{n=1}^{\infty} \frac{(-1)^n A^n (3n-2)!!!}{(3n)!} t^{3n},\ X_2(t) = t + \sum_{n=1}^{\infty} \frac{(-1)^n A^n (3n-1)!!!}{(3n+1)!} t^{3n+1},$$

see [43]. Essentially, the solution $X(t)$ is a linear combination of the fundamental set $\{X_1(t), X_2(t)\}$, where $X_1(t)$ and $X_2(t)$ solve (4.1) with initial conditions $(1, 0)$ and $(0, 1)$, respectively.

If Y_0 and Y_1 are fourth-order random variables, we need to prove the mean fourth convergence of $X_1(t)$ and $X_2(t)$, for each $t \in \mathbb{R}$. That is,

$$\sum_{n=1}^{\infty} \frac{\|A^n\|_4 (3n-2)!!!}{(3n)!} |t|^{3n} < \infty, \quad \sum_{n=1}^{\infty} \frac{\|A^n\|_4 (3n-1)!!!}{(3n+1)!} |t|^{3n+1} < \infty.$$

We work with the first series, as the analysis for the second one is analogous. We have, as a direct consequence of (4.4),

$$\sum_{n=1}^{\infty} \frac{\|A^n\|_4 (3n-2)!!!}{(3n)!} |t|^{3n} \leq \eta \sum_{n=1}^{\infty} \frac{\mathcal{H}^{n-1}(n-1)!^p (3n-2)!!!}{(3n)!} |t|^{3n}.$$

We use d'Alembert's ratio test to derive the radius of convergence of the latter power series. We have

$$\begin{aligned} \lim_{n\to\infty} \frac{\frac{\mathcal{H}^n n!^p (3n+1)!!!}{(3n+3)!} |t|^{3n+3}}{\frac{\mathcal{H}^{n-1}(n-1)!^p (3n-2)!!!}{(3n)!} |t|^{3n}} &= \mathcal{H}|t|^3 \lim_{n\to\infty} \frac{n^p (3n+1)}{(3n+3)(3n+2)(3n+1)} \\ &= \begin{cases} 0, \ 0 \leq p < 2, \\ \frac{\mathcal{H}|t|^3}{9}, \ p = 2. \end{cases} \end{aligned}$$

Then, for $0 \leq p < 2$, the series converges for all $t \in \mathbb{R}$; while for $p = 2$, it converges for $|t| < \sqrt[3]{9/\mathcal{H}}$.

4.3 Random Hermite differential equation

The formal solution to (4.2) is given by a linear combination of a fundamental set $\{X_1(t), X_2(t)\}$,

$$X(t) = Y_0 X_1(t) + Y_1 X_2(t), \tag{4.5}$$

$$X_1(t) = 1 + \sum_{n=0}^{\infty} \frac{t^{2n+2}}{(2n+2)!} \prod_{j=0}^{n} (4j - A), \quad X_2(t) = t + \sum_{n=0}^{\infty} \frac{t^{2n+3}}{(2n+3)!} \prod_{j=0}^{n} (4j + 2 - A), \tag{4.6}$$

see [29]. If Y_0 and Y_1 are fourth-order random variables, we prove that

$$\sum_{n=0}^{\infty} \frac{t^{2n+2}}{(2n+2)!} \left\| \prod_{j=0}^{n} (4j - A) \right\|_4 < \infty, \quad \sum_{n=0}^{\infty} \frac{|t|^{2n+3}}{(2n+3)!} \left\| \prod_{j=0}^{n} (4j + 2 - A) \right\|_4 < \infty,$$

for $t \in \mathbb{R}$. We focus on the convergence analysis for the former series, as the analysis for the second one is analogous.

By Hölder's and the triangular inequalities,

$$\left\|\prod_{j=0}^{n}(4j-A)\right\|_4 \le \prod_{j=0}^{n}\|4j-A\|_{4(n+1)} \le \prod_{j=0}^{n}\left(4j+\|A\|_{4(n+1)}\right) \le \left(4n+\|A\|_{4(n+1)}\right)^{n+1}.$$

The last inequality would be tighter if we used the arithmetic-geometric mean inequality, but the final conclusion that we will derive about the radius of convergence is the same. We have, then,

$$\sum_{n=0}^{\infty}\frac{t^{2n+2}}{(2n+2)!}\left\|\prod_{j=0}^{n}(4j-A)\right\|_4 \le \sum_{n=0}^{\infty}\frac{t^{2n+2}}{(2n+2)!}\left(4n+\|A\|_{4(n+1)}\right)^{n+1}.$$

From (4.4) (just power it to $1/(n+1)$), $\|A\|_{4(n+1)} \le \eta^{1/(n+1)}\mathcal{H}^{n/(n+1)}n!^{p/(n+1)}$. Then

$$\sum_{n=0}^{\infty}\frac{t^{2n+2}}{(2n+2)!}\left\|\prod_{j=0}^{n}(4j-A)\right\|_4 \le \sum_{n=0}^{\infty}\frac{t^{2n+2}}{(2n+2)!}\left(4n+\eta^{1/(n+1)}\mathcal{H}^{n/(n+1)}n!^{p/(n+1)}\right)^{n+1}.$$

Now we use the root test:

$$\lim_{n\to\infty}\sqrt[n]{\frac{t^{2n+2}}{(2n+2)!}\left(4n+\eta^{1/(n+1)}\mathcal{H}^{n/(n+1)}n!^{p/(n+1)}\right)^{n+1}}$$
$$= \lim_{n\to\infty}\frac{|t|^{2+\frac{2}{n}}}{(2n+2)!^{\frac{1}{n}}}\left(4n+\eta^{1/(n+1)}\mathcal{H}^{n/(n+1)}n!^{p/(n+1)}\right)^{\frac{n+1}{n}} = \begin{cases}0, & 0\le p<2,\\ t^2\frac{\mathcal{H}}{4}, & p=2.\end{cases}$$

As a consequence, for $0 \le p < 2$ the series converges for all $t \in \mathbb{R}$; for $p = 2$ it converges for $|t| < 2/\sqrt{\mathcal{H}}$.

4.4 Random Laguerre differential equation

The formal solution to (4.3) is expressed as

$$X(t) = Y_0\sum_{n=0}^{\infty}t^n\prod_{k=1}^{n}\frac{k-1-A}{k^2}, \tag{4.7}$$

see [44]. In this case, there is only one initial condition because 0 is a regular singular point. If Y_0 is a fourth-order random variable, we prove that

$$\sum_{n=0}^{\infty} |t|^n \left\| \prod_{k=1}^{n} \frac{k-1-A}{k^2} \right\|_4 < \infty.$$

By Hölder's and the triangular inequalities,

$$\begin{aligned} \left\| \prod_{k=1}^{n} \frac{k-1-A}{k^2} \right\|_4 &\leq \prod_{k=1}^{n} \left\| \frac{k-1-A}{k^2} \right\|_{4n} \leq \prod_{k=1}^{n} \frac{k+\|A\|_{4n}}{k^2} \\ &\leq \frac{\left(n+\eta^{1/n}\mathcal{H}^{n/(n-1)}(n-1)!^{p/n}\right)^n}{n!^2}. \end{aligned}$$

At this point, we would like to point out a mistake in the inequality derived for $\| \prod_{k=1}^{n} \frac{k-1-A}{k^2} \|_4$ in [44], as the numerator of the final bound found by the authors should be powered to the n-th. This mistake, however, does not change their conclusions about the radius of convergence.

We use the root test:

$$\begin{aligned} &\lim_{n\to\infty} \sqrt[n]{\frac{\left(n+\eta^{1/n}\mathcal{H}^{n/(n-1)}(n-1)!^{p/n}\right)^n}{n!^2}|t|^n} \\ &= |t| \lim_{n\to\infty} \frac{n+\eta^{1/n}\mathcal{H}^{n/(n-1)}(n-1)!^{p/n}}{n!^{2/n}} = \begin{cases} 0, \ 0 \leq p < 2, \\ |t|\mathcal{H}, \ p = 2. \end{cases} \end{aligned}$$

The series thus converges on $\mathbb{R}$ when $0 \leq p < 2$, and on $(-1/\mathcal{H}, 1/\mathcal{H})$ when $p = 2$.

Remark 4.1 If the random vector (Y_0, Y_1) and A are independent, then we can relax the hypotheses to Y_0 and Y_1 second-order random variables and

$$\|A^n\|_2 \leq \eta\mathcal{H}^{n-1}(n-1)!^p.$$

Indeed, this is because $\|UV\|_2 = \|U\|_2\|V\|_2$ whenever U and V are two independent random variables.

Remark 4.2 In the contributions [43, 29, 44], when A is unbounded, its support gets truncated. Suppose that Y_0 and Y_1 are mean fourth integrable and A satisfies (4.4). Consider truncations $A^{(m)} = A1_{\{|A|\leq a_m\}}$, where $\lim_{m\to\infty} a_m = \infty$. These truncations are bounded and satisfy $\lim_{m\to\infty} \|A^{(m)} - A\|_k = 0$, for

all $1 \leq k < \infty$ (by the dominated convergence theorem). Let $X^{(m)}(t)$ be the solution to the equation with coefficient $A^{(m)}$, and let $X(t)$ be the solution when the coefficient is A. These solutions are given by random power series, see the previous sections. We have that $\lim_{m\to\infty} \|X^{(m)}(t) - X(t)\|_2 = 0$, for each $t \in \mathbb{R}$ (this is a consequence of the dominated convergence theorem for series).

4.5 About the hypotheses

As a consequence of Stirling's formula, $(n-1)! \approx \sqrt{2\pi(n-1)}(\frac{n-1}{\mathrm{e}})^{n-1}$ as $n \to \infty$, hypothesis (4.4) is equivalent to

$$\|A^n\|_4 \leq \gamma \mathcal{C}^{n-1}(n-1)^{p(n-1)} \tag{4.8}$$

for $n \geq n_1$, for constants $n_1, \gamma, \mathcal{C}, p > 0$.

This condition (4.8) might be easier to check in practice. For example, let us prove that the Poisson(λ) distribution for A satisfies (4.4), by taking advantage of the well-known convergence of the Binomial distribution to the Poisson distribution under specific hypotheses on their corresponding parameters. Let $k \geq 1$. Let $V_{s,\alpha} \sim$ Binomial(s, α), where $s \geq k$ is a positive integer and $\alpha \in (0, 1)$. By [92, Prop. 4.7], there is a constant $C > 0$ independent of α, s and k such that

$$\|V_{s,\alpha}\|_k \leq C \frac{k}{\log(\frac{k}{\alpha s})},$$

when $k/(\alpha s) \geq \mathrm{e}$. Let $s \to \infty$, $\alpha \to 0$, $\alpha s \to \lambda$. Then $V_{s,\alpha}$ converges to a Poisson(λ) random variable U, which satisfies

$$\|U\|_k \leq C \frac{k}{\log(\frac{k}{\lambda})}$$

when $k \geq \mathrm{e}\lambda$. In particular, if $A \sim$ Poisson(λ), then

$$\|A^n\|_4 = \|A\|_{4n}^n \leq C^n \frac{(4n)^n}{\log^n(4n/\lambda)} \leq C^n (4n)^n$$

for large n, so (4.8) holds with $p = 1$.

Up to now, we know that (4.4) is fulfilled by the bounded, Normal, Gamma and Poisson distributions, with $p = 0$, $p = 1/2$ and $p = 1$, respectively. We show that condition (4.4) for $p \leq 1$ is equivalent to $\phi_A(t) < \infty$ in a neighborhood of 0, where $\phi_A(t) = \mathbb{E}[\mathrm{e}^{tA}]$ denotes the moment-generating function of A. We

use [116, Th. A, p. 5], (a)⇔(c): $\phi_A(t) < \infty$ in a neighborhood of 0 if and only if $\mathbb{E}[A^{4k}] \leq C^k(4k)!$ for certain $C > 0$. Now, $C^k(4k)! \sim D^{4k}k^{4k}$, for certain $D > 0$, by Stirling's approximation. Then $\mathbb{E}[A^{4k}] \leq C^k(4k)!$ is equivalent to $\|A^k\|_4 \leq D^k k^k$, which is in turn equivalent to (4.8) with $p = 1$.

Condition (4.4) for $p \leq 2$ and $A \geq 0$ is equivalent to $\phi_{\sqrt{A}}(t) < \infty$ in a neighborhood of 0, by [116, Th. B, p. 6], (a)⇔(c).

Example 4.3 Condition (4.4) is not satisfied by the Log-Normal distribution, for any $0 \leq p \leq 2$. If A follows a Log-Normal distribution of parameters $\mu \in \mathbb{R}$ and $\sigma > 0$, then $\mathbb{E}[A^n] = \mathrm{e}^{n\mu+n^2\sigma^2/2}$. In particular, $\|A^n\|_4 = \mathrm{e}^{n\mu+2n^2\sigma^2}$. This quantity grows faster than $\mathcal{C}^{n-1}(n-1)^{p(n-1)}$ as $n \to \infty$, for any $\mathcal{C}, p > 0$, as the limit of their ratio is infinity. As a conclusion, the random Fröbenius method does not work with the Log-Normal distribution. This issue also occurs with other methods for uncertainty quantification, namely Monte Carlo simulation and polynomial chaos expansions [172]. Due to the large growth of the moments of the Log-Normal distribution and its fat tails, the classical Monte Carlo procedure does not work well with this distribution. On the other hand, polynomial chaos expansions and stochastic Galerkin projections do not converge for the Log-Normal distribution [66].

Example 4.4 Consider Laguerre's equation (4.3), $A \sim \text{Laplace}(-0.5, 0.5)$ ($\mu = -0.5$ is the location parameter and $b = 0.5$ is the scale parameter) and $Y_0 \sim \text{Uniform}(0.1, 0.2)$. These two random variables are independent. The moment-generating function of the Laplace distribution is finite in a neighborhood of zero defined by the interval $(-1/b, 1/b)$. Then, according to the theoretical discussion, there exists a mean square solution $X(t)$ defined by (4.7), $t \in \mathbb{R}$. By taking the N-th partial sum of the random power series, the expectation and the variance of $X(t)$, $\mathbb{E}[X(t)]$ and $\mathbb{V}[X(t)]$, can be approximated as $N \to \infty$. See the formulae from [44, p. 289] (the idea relies on using the linearity of the expectation and the precomputed moments of A). The convergence is exponentially fast with N, but not uniformly in t; as t increases, a larger order N is required. In Tables 4.1 and 4.2, we show the approximations of $\mathbb{E}[X(t)]$ and $\mathbb{V}[X(t)]$ for different N's. The results are compared against Monte Carlo simulation with 100,000 realizations. While the Fröbenius method gives correct significant digits very fast, the Monte Carlo procedure shows much slower convergence.

Example 4.5 Consider Hermite's equation (4.2), with inputs $A \sim \text{Poisson}(2)$ (the parameter 2 is the mean), $Y_0 \sim \text{Uniform}(0.1, 0.2)$ and $Y_1 = -1$, where A

t	$N=5$	$N=6$	$N=7$	$N=8$	$N=9$	$N=10$	MC
0	0.150000	0.150000	0.150000	0.150000	0.150000	0.150000	0.149982
0.2	0.167021	0.167021	0.167021	0.167021	0.167021	0.167021	0.167247
0.4	0.188737	0.188738	0.188738	0.188738	0.188738	0.188738	0.188417
0.6	0.216310	0.216320	0.216321	0.216321	0.216321	0.216321	0.215913
0.8	0.251174	0.251230	0.251238	0.251239	0.251239	0.251239	0.250905
1.0	0.295083	0.295298	0.295333	0.295338	0.295339	0.295339	0.295943
1.2	0.350149	0.350791	0.350917	0.350939	0.350943	0.350943	0.35014
1.4	0.418889	0.420507	0.420879	0.420954	0.420968	0.420970	0.421562

Table 4.1: Approximations of the expectation of the solution (4.7) to the random Laguerre differential equation (4.3) using the Fröbenius method, for different orders of truncation N and Monte Carlo (MC) simulation. Example 4.4.

t	$N=5$	$N=6$	$N=7$	$N=8$	$N=9$	$N=10$	MC
0	0.000833333	0.000833333	0.000833333	0.000833333	0.000833333	0.000833333	0.000834754
0.2	0.00161217	0.00161217	0.00161217	0.00161217	0.00161217	0.00161217	0.00161509
0.4	0.00428272	0.00428300	0.00428302	0.00428302	0.00428302	0.00428302	0.00427721
0.6	0.0105189	0.0105246	0.0105253	0.0105253	0.0105253	0.0105253	0.0104690
0.8	0.0234742	0.0235277	0.0235358	0.0235368	0.0235370	0.0235370	0.0235577
1.0	0.0489538	0.0492733	0.0493346	0.0493448	0.0493463	0.0493465	0.0486700
1.2	0.0973951	0.0988329	0.0991692	0.0992369	0.0992489	0.0992507	0.0991242
1.4	0.187108	0.192424	0.193900	0.194250	0.194323	0.194000	0.189003

Table 4.2: Approximations of the variance of the solution (4.7) to the random Laguerre differential equation (4.3) using the Fröbenius method, for different orders of truncation N and Monte Carlo (MC) simulation. Example 4.4.

and Y_0 are assumed to be independent. The moment-generating function of the Poisson distribution is finite on the whole $\mathbb{R}$, therefore there exists a mean square solution $X(t)$ defined by (4.5)–(4.6), $t \in \mathbb{R}$. By using the formulae from [29, Section 6], based on truncating the power series to the N-th partial sum, the expectation and the variance of $X(t)$, $\mathbb{E}[X(t)]$ and $\mathbb{V}[X(t)]$, can be approximated as $N \to \infty$. In Tables 4.3 and 4.4, we show the approximations of $\mathbb{E}[X(t)]$ and $\mathbb{V}[X(t)]$ for different N's, as well as the Monte Carlo estimates with $100,000$ realizations.

t	$N=5$	$N=6$	$N=7$	$N=8$	$N=9$	$N=10$	MC
0	0.150000	0.150000	0.150000	0.150000	0.150000	0.150000	0.149969
0.2	-0.0560256	-0.0560256	-0.0560256	-0.0560256	-0.0560256	-0.0560256	-0.0559303
0.4	-0.274510	-0.274510	-0.274510	-0.274510	-0.274510	-0.274510	-0.274442
0.6	-0.507193	-0.507193	-0.507193	-0.507193	-0.507193	-0.507193	-0.507183
0.8	-0.758494	-0.758494	-0.758494	-0.758494	-0.758494	-0.758494	-0.758386
1.0	-1.03813	-1.03814	-1.03814	-1.03814	-1.03814	-1.03814	-1.03799
1.2	-1.36655	-1.3666	-1.36661	-1.36661	-1.36661	-1.36661	-1.36748
1.4	-1.78648	-1.78703	-1.78715	-1.78718	-1.78718	-1.78718	-1.79147

Table 4.3: Approximations of the expectation of the solution (4.5)–(4.6) to the random Hermite differential equation (4.2) using the Fröbenius method, for different orders of truncation N and Monte Carlo (MC) simulation. Example 4.5.

t	$N=5$	$N=6$	$N=7$	$N=8$	$N=9$	$N=10$	MC
0	0.000833333	0.000833333	0.000833333	0.000833333	0.000833333	0.000833333	0.000831451
0.2	0.000773883	0.000773883	0.000773883	0.000773883	0.000773883	0.000773883	0.000776030
0.4	0.000596974	0.000596974	0.000596974	0.000596974	0.000596974	0.000596974	0.000597567
0.6	0.000665335	0.000665336	0.000665336	0.000665336	0.000665336	0.000665336	0.000664739
0.8	0.00514877	0.00514884	0.00514884	0.00514884	0.00514884	0.00514884	0.00519014
1.0	0.0343588	, 0.0343611	0.0343612	0.0343613	0.0343613	0.0343613	0.0345285
1.2	0.163114	0.163173	0.163181	0.163182	0.163182	0.163182	0.163717
1.4	0.645706	0.646805	0.647026	0.647070	0.647077	0.647079	0.647087

Table 4.4: Approximations of the variance of the solution (4.5)–(4.6) to the random Hermite differential equation (4.2) using the Fröbenius method, for different orders of truncation N and Monte Carlo (MC) simulation. Example 4.5.

Example 4.6 In this last example, we consider Hermite's equation (4.2), with coefficient $A \sim \text{Weibull}(a, b)$ (a is the scale parameter and b is the shape parameter) and initial conditions $Y_0 \sim \text{Uniform}(0.1, 0.2)$ and $Y_1 = -1$, where A and Y_0 are assumed to be independent. The moments of the Weibull distribution are well-known: $\mathbb{E}[A^m] = a^m\Gamma(1 + m/b)$, where Γ is the Gamma function. Using this formula, one can estimate the ratio $\|A^{m+1}\|_2/\|A^m\|_2$ (we

use 2-norms instead of 4-norms because of the independence, see Remark 4.1). By Stirling's formula, $\Gamma(x+1) \sim \sqrt{2\pi x}(x/\mathrm{e})^x$ as $x \to \infty$, it is easy to obtain, by direct computations, that

$$\frac{\|A^{m+1}\|_2}{\|A^m\|_2} \sim a \left(\frac{2}{b}\right)^{\frac{1}{b}} m^{\frac{1}{b}}.$$

That is, $p = 1/b$, $\mathcal{H} = a(2/b)^{1/b}$ and $\eta = \|A\|_2 = a\sqrt{\Gamma(1+2/b)}$. We analyze the mean square convergence of the series defined by (4.5)–(4.6). We show the approximations of the variance of $X(t)$, $\mathbb{V}[X(t)]$, for orders of truncation $N = 25, 26, 27$, scale parameter $a = 1$ and shape parameters $b = 2/3, 1/2, 1/3$. Notice that, for $b = 2/3$, we have $p = 3/2 < 2$, so convergence on the whole real line is expected by the theoretical results. For $b = 1/2$, we have that p equals the threshold 2 for convergence: the series defined by (4.5)–(4.6) only converges in a small interval around 0, given by $(-2/\sqrt{\mathcal{H}}, 2/\sqrt{\mathcal{H}}) = (-1/2, 1/2)$. Finally, for $b = 1/3$, we have $p = 3 > 2$, therefore the series given by (4.5)–(4.6) is not expected to converge for any $t \neq 0$. The numerical results are presented in Tables 4.5, 4.6 and 4.7. They agree with our theoretical discussion about the convergence domain. The results of the Monte Carlo simulation using $100,000$ realizations are also shown for validation.

t	$N = 25$	$N = 26$	$N = 27$	MC
0	0.000833333	0.000833333	0.000833333	0.000830033
0.2	0.000802005	0.000802005	0.000802005	0.000813572
0.4	0.000687696	0.000687696	0.000687696	0.000669479
0.6	0.00122022	0.00122022	0.00122022	0.0012146
0.8	0.00975692	0.00975692	0.00975692	0.0103762
1.0	0.0559536	0.0559536	0.0559536	0.0541698
1.2	0.233081	0.233081	0.233081	0.233012
1.4	0.830763	0.830763	0.830763	0.827578

Table 4.5: Approximations of the variance of the solution (4.5)–(4.6) to the random Hermite differential equation (4.2) using the Fröbenius method, for $b = 2/3$, different orders of truncation N and Monte Carlo (MC) simulation. Example 4.6.

t	$N = 25$	$N = 26$	$N = 27$	MC
0	0.000833333	0.000833333	0.000833333	0.000838800
0.2	0.000818187	0.000818187	0.000818187	0.000824624
0.4	0.000691749	0.000691749	0.000691749	0.000690404
0.6	238750	489177	1.00245×10^{6}	0.00361506
0.8	1.05641×10^{18}	6.85769×10^{18}	4.45158×10^{19}	0.0275597
1.0	6.82155×10^{27}	1.08386×10^{29}	1.72173×10^{30}	0.121455

Table 4.6: Approximations of the variance of the solution (4.5)–(4.6) to the random Hermite differential equation (4.2) using the Fröbenius method, for $b = 1/2$, different orders of truncation N and Monte Carlo (MC) simulation. Example 4.6.

t	$N = 25$	$N = 26$	$N = 27$	MC
0	0.000833333	0.000833333	0.000833333	0.000827802
0.1	3.67222×10^{32}	6.87049×10^{34}	1.38863×10^{37}	0.000927809
0.2	4.95578×10^{62}	1.48136×10^{66}	4.78402×10^{69}	0.00105165
0.3	2.05526×10^{80}	3.11183×10^{84}	5.09011×10^{88}	0.00104901

Table 4.7: Approximations of the variance of the solution (4.5)–(4.6) to the random Hermite differential equation (4.2) using the Fröbenius method, for $b = 1/3$, different orders of truncation N and Monte Carlo (MC) simulation. Example 4.6.

4.6 Conclusions and perspectives

In this chapter we have studied Airy, Hermite and Laguerre differential equations with random inputs. Using the Fröbenius method, the mean square stochastic solution has been constructed in the form of a power series. The main goal has been to weaken the usual hypothesis of boundedness for the equation coefficient, so that the class of its probability distributions is enlarged to those having certain growth of the moments. In particular, we include the probability distributions having finite moment-generating function around the origin.

More research on these methods should be carried out in the future. For instance, the solution to the random Legendre differential equation still requires boundedness of the equation coefficient, see Chapter 2. In the general case of second-order linear differential equations, see Chapter 3, all the equation coefficients are taken as bounded random variables. It would be interesting to generalize the theoretical discussion therein to allow unbounded distributions.

Acknowledgements

This work has been supported by the Spanish Ministerio de Economía y Competitividad grant MTM2017–89664–P. I acknowledge the doctorate scholarship granted by Programa de Ayudas de Investigación y Desarrollo (PAID), Universitat Politècnica de València.

The main results of this chapter have been submitted for publication.

Chapter 5

Random second-order linear differential equations via adapted gPC: a comparative study with the random Fröbenius method and Monte Carlo simulation

This chapter presents a methodology to quantify computationally the uncertainty of random non-autonomous second-order linear differential equations, via adaptive generalized Polynomial Chaos (gPC) and the stochastic Galerkin projection technique. Unlike the random Fröbenius method, which can only deal with particular random linear differential equations and needs the random inputs (coefficients and forcing term) to be analytic, adaptive gPC allows approximating the expectation and covariance of the solution stochastic process to general random second-order linear differential equations. The random inputs are allowed to functionally depend on random variables that may be independent or dependent, both absolutely continuous or discrete with infinitely many point masses. These hypotheses include a wide variety of particular differential equations, which might not be solvable via the random Fröbenius method, in which the random input coefficients may be expressed via a Karhunen-Loève expansion.

5.1 Introduction and Preliminaries

A powerful tool to deal with random differential equations is generalized Polynomial Chaos (gPC) expansions [172, 173]. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space. We will work in the Hilbert space $(\mathrm{L}^2(\Omega), \langle \cdot, \cdot \rangle)$ that consists of second-order random variables, i.e., random variables with finite variance, where the inner product is defined by $\langle \zeta_1, \zeta_2 \rangle = \mathbb{E}[\zeta_1 \zeta_2]$, being $\mathbb{E}[\cdot]$ the expectation operator. In its classical formulation, gPC consists in writing a random vector $\zeta : \Omega \to \mathbb{R}^n$ as a limit of multivariate polynomials evaluated at a random vector $Z : \Omega \to \mathbb{R}^n$: $\zeta \approx \sum_{i=0}^{P} \hat{\zeta}_i \phi_i(Z)$. Here $\{\phi_i(Z)\}_{i=0}^{\infty}$ is a sequence of orthogonal polynomials in Z: $\mathbb{E}[\phi_i(Z)\phi_j(Z)] = \int_{\mathbb{R}^n} \phi_i(z)\phi_j(z)\, \mathrm{d}\mathbb{P}_Z(z) = \gamma_i \delta_{ij}$, where $\mathbb{P}_Z = \mathbb{P} \circ Z^{-1}$ is the law of Z and δ_{ij} is the Kronecker delta symbol. A stochastic Galerkin method can be applied to approximate the solution to random differential equations [172, Ch. 6]. For some applications of this theory, see for example [40, 171]. Given the random vector Z, the sequence $\{\phi_i(Z)\}_{i=0}^{\infty}$ of orthogonal polynomials is taken from the Askey-Wiener scheme of hypergeometric orthogonal polynomials, by taking into account the density function f_Z of Z (if Z is absolutely continuous) or the discrete masses of Z (if Z is discrete), [172, 173].

Chapter 3 does not compare the Fröbenius method with gPC expansions, although it has been proved to be a powerful technique to deal with general continuous and discrete stochastic systems. Due to the spectral mean square convergence of the Galerkin projections, the expectation and the variance statistics of the response process can be approximated with small orders of truncation. In the particular setting of random second-order linear differential equations, only [35, 36, 50] analyze the application of gPC expansions to Airy's random differential equation. Paper [35] analyzes the best trial basis for applying gPC expansions; reference [36] compares gPC and random power series expansions when approximating statistical moments; and [50] studies the application of gPC expansions for dependent Gaussian uncertainty, by mapping the vector of input coefficients to independent Gaussian random variables, and compares it with the random Fröbenius method. It may be interesting to analyze the application of gPC expansions to general random second-order linear differential equations.

In the recent articles [34, 49, 51], an adaptive gPC method has been developed to approximate the solutions of random differential equations. Instead of taking the orthogonal polynomials from the Askey-Wiener scheme, the authors construct them directly from the random inputs that are involved in the corresponding random differential equation formulation.

More explicitly, in [34], the authors considered the random differential equation $F(t, y, \dot{y}) = 0$, $y(t_0) = y_0$, where $F : \mathbb{R}^{2q+1} \to \mathbb{R}^q$ and $y(t) = (y^1(t), \ldots, y^q(t))^\top$, where $\top$ denotes the transpose operator. The set $\{\zeta_1, \ldots, \zeta_s\}$ represents independent and absolutely continuous random input parameters in the random differential equation problem.

For each $1 \leq i \leq s$, it is considered the canonical basis of polynomials in ζ_i of degree at most p: $\mathcal{C}_i^p = \{1, \zeta_i, (\zeta_i)^2, \ldots, (\zeta_i)^p\}$. One defines the following inner product, with weight function given by the density of ζ_i: $\langle g(\zeta_i), h(\zeta_i)\rangle_{\zeta_i} = \int_{\mathbb{R}} g(\zeta_i)h(\zeta_i)f_{\zeta_i}(\zeta_i)\,\mathrm{d}\zeta_i$. Using a Gram-Schmidt orthonormalization procedure, one obtains a sequence of orthonormal polynomials in ζ_i with respect to $\langle , \rangle_{\zeta_i}$: $\Xi_i^p = \{\phi_0^i(\zeta_i), \ldots, \phi_p^i(\zeta_i)\}$. The authors build a sequence of orthonormal multivariate polynomials in $\zeta = (\zeta_1, \ldots, \zeta_s)^\top$ of degree at most p with respect to the inner product $\langle g(\zeta), h(\zeta)\rangle_\zeta = \int_{\mathbb{R}^s} g(\zeta)h(\zeta)f_\zeta(\zeta)\,\mathrm{d}\zeta$. To do so, they build the simple tensor product $\phi_j(\zeta) = \phi_{p_1}^1(\zeta_1)\cdots\phi_{p_s}^s(\zeta_s)$, $1 \leq j \leq P$, where j is associated in a bijective manner to the multi-index $(p_1, \ldots, p_s)$ in such a way that 1 corresponds to $(0, \ldots, 0)$ (for example, a graded lexicographic ordering [172, p. 66]) and $P = (p+s)!/(p!s!)$. By the independence between $\zeta_1, \ldots, \zeta_s$, the built sequence $\Xi = \{\phi_j(\zeta)\}_{j=1}^P$ is orthonormal with respect to $\langle , \rangle_\zeta$.

Once the basis is constructed, one looks for an approximate solution $y(t) \approx \sum_{j=1}^P y_j(t)\phi_j(\zeta)$. Then, $F(t, \sum_{j=1}^P y_j(t)\phi_j(\zeta), \sum_{j=1}^P \dot{y}_j(t)\phi_j(\zeta)) = 0$. To obtain the deterministic coefficients $y_j(t)$, one computes the inner products $\langle F(t, \sum_{j=1}^P y_j(t)\phi_j(\zeta), \sum_{j=1}^P \dot{y}_j(t)\phi_j(\zeta)), \phi_k(\zeta)\rangle_\zeta = 0$, $k = 1, \ldots, P$. In this manner, one arrives at a deterministic system of P differential equations, which may be solved by standard numerical techniques. Once $y_1(t), \ldots, y_P(t)$ have been computed, the expectation of the actual solution $y(t)$ is approximated by $y_1(t)$ and the covariance matrix is approximated by $\sum_{i=1}^P y_i(t)y_i(t)^\top$.

In [51], the authors use the Random Variable Transformation technique [47, Th. 1] in case that some random input parameters appearing in the random differential equation come from mappings of absolutely continuous random variables, whose probability density function is known.

In [49], the authors focus on the case that the random inputs $\zeta_1, \ldots, \zeta_s$ are not independent. They consider the canonical bases $\mathcal{C}_i^p = \{1, \zeta_i, (\zeta_i)^2, \ldots, (\zeta_i)^p\}$, for $1 \leq i \leq s$, and construct a sequence of multivariate polynomials in ζ, via a simple tensor product: $\phi_j(\zeta) = \zeta_1^{p_1}\cdots\zeta_s^{p_s}$, where $1 \leq j \leq P$ corresponds to the multi-index $(p_1, \ldots, p_s)$ and $P = (p+s)!/(p!s!)$. Notice that this new sequence $\{\phi_j(\zeta)\}_{j=1}^P$ is not orthonormal with respect to $\langle , \rangle_\zeta$. However, one

proceeds with the random differential equation problem as in [34] and, in practice, one obtains good approximations of the expectation and covariance of $y(t)$.

Based on ample numerical evidence, the gPC-based methods described in [34, 49, 51, 172, 173] converge in the mean square sense at spectral rate. Some theoretical results that justify this assertion are presented in [172, pp. 33–35, p. 73], [79, Th. 2.2], [18, 66, 150, 151].

In this chapter we deal with general random non-autonomous second-order linear differential equations:

$$\begin{cases} \ddot{X}(t) + A(t)\dot{X}(t) + B(t)X(t) = C(t),\ t \in \mathbb{R}, \\ X(t_0) = Y_0, \\ \dot{X}(t_0) = Y_1. \end{cases} \tag{5.1}$$

Our goal is to obtain approximations of the stochastic solution $X(t)$ as well as of its main statistical features, by taking advantage of the adaptive gPC techniques [34, 49]. Here, $A(t)$, $B(t)$ and $C(t)$ are stochastic processes and Y_0 and Y_1 are random variables in an underlying complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The term $X(t)$ is the solution stochastic process to the random initial value problem (5.1) in some probabilistic sense. We will detail conditions for existence and uniqueness of solution in the following section.

In this chapter, we want to go one step further and we will perform a computational analysis based upon adaptive gPC, by showing its capability to deal with the general random initial value problem (5.1) that comprises Airy, Hermite, Legendre, Laguerre and Bessel differential equations, or any other formulation of (5.1) based on analytic data processes, just as particular cases.

The chapter is organized as follows. Section 5.2 describes the application of adaptive gPC to solve the random initial value problem (5.1) and the computation of the expectation and covariance of $X(t)$. The study is split into two cases depending on the probabilistic dependence of the random inputs. In Section 5.3, we show the algorithms corresponding to the theory previously developed in Section 5.2. Section 5.4 is addressed to show particular examples of (5.1) where adaptive gPC, Fröbenius method and Monte Carlo simulation are carried out to obtain approximations for the expectation, variance and covariance of the solution stochastic process. It is evinced that adaptive gPC provides the same results as the Fröbenius method with small orders of basis p, and, moreover, in cases where the Fröbenius method is not applicable, adaptive gPC might be successful. Finally, in Section 5.5, conclusions are drawn.

5.2 Method

Consider the random initial value problem (5.1), where

$$A(t) = a_0(t) + \sum_{i=1}^{d_A} a_i(t)\gamma_i,\ B(t) = b_0(t) + \sum_{i=1}^{d_B} b_i(t)\eta_i,\ C(t) = c_0(t) + \sum_{i=1}^{d_C} c_i(t)\xi_i, \tag{5.2}$$

being $\gamma_1, \ldots, \gamma_{d_A}$, $\eta_1, \ldots, \eta_{d_B}$ and $\xi_1, \ldots, \xi_{d_C}$ random variables (not necessarily independent) and $a_0(t), \ldots, a_{d_A}(t)$, $b_0(t), \ldots, b_{d_B}(t)$ and $c_0(t), \ldots, c_{d_C}(t)$ real functions. Representation (5.2) for the input stochastic processes includes truncated random power series [160, p. 99] and Karhunen-Loève expansions [172, Ch. 4], [119, Ch. 5]. This is an improvement with respect to the random Fröbenius method, in which $A(t)$, $B(t)$ and $C(t)$ are only expressed as random power series.

As we are interested in constructive computational aspects of uncertainty quantification, we will assume that there exists a unique solution stochastic process $X(t)$ to initial value problem (5.1) in some probabilistic sense, for instance, sample-path [161, SP problem] [160, Appendix A], or $\mathrm{L}^q(\Omega)$ sense [160], in such a way that $\mathbb{E}[X(t)^2] < \infty$ for each t. We detail the conditions under which there exists a unique solution $X(t)$ to (5.1) in the following propositions. Proposition 5.1, which is concerned with sample-path solutions, is a direct consequence of the deterministic theory on ordinary differential equations (Carathéodory theory on the existence of absolutely continuous solutions [81, pp. 28–30]). Proposition 5.2 takes advantage of a natural generalization to $\mathrm{L}^q(\Omega)$ random calculus of the classical Picard's theorem for deterministic ordinary differential equations [160, Th. 5.1.2], see Theorem 3.5 from Chapter 3.

Proposition 5.1 (Sample-path solution) *If $A(t)$, $B(t)$ and $C(t)$ have real integrable sample paths, then there exists a unique solution stochastic process $X(t)$ to (5.1) with C^1 sample paths and derivative $\dot{X}(t)$ with absolutely continuous sample paths (i.e., $X(t)$ is a classical solution that belongs to the Sobolev space $W^{2,1}$). Moreover, if $A(t)$, $B(t)$ and $C(t)$ have continuous sample paths, then $X(t)$ has C^2 sample paths.*

Proposition 5.2 (L^q-solution) *If $A(t)$ and $B(t)$ are continuous stochastic processes in the $\mathrm{L}^\infty(\Omega)$ sense, and the source term $C(t)$ is continuous in the $\mathrm{L}^q(\Omega)$ setting, then there exists a unique solution $X(t)$ to (5.1) in the $\mathrm{L}^q(\Omega)$ sense.*

Our goal is to approximate the solution stochastic process $X(t)$ to the random initial value problem (5.1) by using adaptive gPC, which is described in [34, 49] and has been reviewed in Section 5.1. In the case that the random inputs $\gamma_1, \ldots, \gamma_{d_A}$, $\eta_1, \ldots, \eta_{d_B}$, $\xi_1, \ldots, \xi_{d_C}$, Y_0 and Y_1 are independent, we will use the method from [34], whereas in the case that they are not independent, [49] will be utilized. In [34, 49], the random inputs are assumed to be absolutely continuous, so that the weights in the inner products are given by density functions. Notice, however, that a discrete distribution with infinitely many point masses can be given to the random inputs. Indeed, the corresponding inner product becomes an integral with respect to a discrete law, which is a series with weights being the probabilities of the point masses. Moreover, since the support has infinite cardinality, the corresponding canonical basis of polynomials has infinite dimension, so that its length p can grow up to infinity.

For ease of notation and to identify the notation with the one used in Section 5.1, we denote the random inputs $\gamma_1, \ldots, \gamma_{d_A}$, $\eta_1, \ldots, \eta_{d_B}$, $\xi_1, \ldots, \xi_{d_C}$, Y_0 and Y_1 as $\zeta_1, \ldots, \zeta_s$, where $s = d_A + d_B + d_C + 2$. The random variables $\zeta_1, \ldots, \zeta_s$ are not necessarily independent, and they are absolutely continuous or discrete random variables with infinitely many point masses. We will denote $\zeta = (\zeta_1, \ldots, \zeta_s)^\top$. The space of polynomials evaluated at ζ_i of degree at most p will be denoted by $\mathcal{P}_p[\zeta_i]$. The space of multivariate polynomials evaluated at ζ of degree at most P will be written as $\mathcal{P}_P^s[\zeta]$.

In the next development, we distinguish two cases depending on whether the random inputs $\zeta_1, \ldots, \zeta_s$ are independent or not.

5.2.1 The random inputs are independent

In the notation from [34] and Section 5.1, let $\mathcal{C}_i^p = \{1, \zeta_i, \ldots, \zeta_i^p\}$ be the canonical basis of $\mathcal{P}_p[\zeta_i]$, for $i = 1, \ldots, s$. Let $\Xi_i^p = \{\phi_0^i(\zeta_i), \ldots, \phi_p^i(\zeta_i)\}$ be the orthonormalization of $\mathcal{C}_i^p$ with respect to the inner product defined by the law $\mathbb{P}_{\zeta_i}$, via a Gram-Schmidt procedure. Let $\Xi = \{\phi_1(\zeta), \ldots, \phi_P(\zeta)\}$ be the orthonormal basis of $\mathcal{P}_P^s[\zeta]$ with respect to the law $\mathbb{P}_\zeta = \mathbb{P}_{\zeta_1} \times \cdots \times \mathbb{P}_{\zeta_s}$, where $P = (p+s)!/(p!s!)$.

We approximate the solution stochastic process $X(t) \approx \sum_{i=1}^P x_i(t)\phi_i(\zeta)$ by imposing the right-hand side to be a solution to random initial value prob-

lem (5.1):

$$\sum_{i=1}^{P} \ddot{x}_i(t)\phi_i(\zeta) + \left(a_0(t) + \sum_{i=1}^{d_A} a_i(t)\gamma_i\right)\left(\sum_{i=1}^{P} \dot{x}_i(t)\phi_i(\zeta)\right) + \left(b_0(t) + \sum_{i=1}^{d_B} b_i(t)\eta_i\right)\left(\sum_{i=1}^{P} x_i(t)\phi_i(\zeta)\right) = c_0(t) + \sum_{i=1}^{d_C} c_i(t)\xi_i. \tag{5.3}$$

We apply the stochastic Galerkin projection technique. By multiplying by $\phi_k(\zeta)$, $k = 1, \ldots, P$, applying expectations, using the orthonormality of Ξ and the fact that $\phi_1 = 1$, we obtain:

$$\ddot{x}_k(t) + a_0(t)\dot{x}_k(t) + \sum_{i=1}^{d_A}\sum_{j=1}^{P} a_i(t)\dot{x}_j(t)\mathbb{E}[\gamma_i\phi_j(\zeta)\phi_k(\zeta)] + b_0(t)x_k(t) + \sum_{i=1}^{d_B}\sum_{j=1}^{P} b_i(t)x_j(t)\mathbb{E}[\eta_i\phi_j(\zeta)\phi_k(\zeta)] = c_0(t)\delta_{1k} + \sum_{i=1}^{d_C} c_i(t)\mathbb{E}[\xi_i\phi_k(\zeta)]. \tag{5.4}$$

Let us put this equation in matrix form. Consider the $P \times P$ matrices M and N defined by

$$M_{kj}(t) = \sum_{i=1}^{d_A} a_i(t)\mathbb{E}[\gamma_i\phi_j(\zeta)\phi_k(\zeta)],\ N_{kj}(t) = \sum_{i=1}^{d_B} b_i(t)\mathbb{E}[\eta_i\phi_j(\zeta)\phi_k(\zeta)], \tag{5.5}$$

for $k, j = 1, \ldots, P$. Consider the vector q of length P with

$$q_k = \sum_{i=1}^{d_C} c_i(t)\mathbb{E}[\xi_i\phi_k(\zeta)], \tag{5.6}$$

for $k = 1 \ldots, P$. We rewrite (5.4) as a deterministic system of P differential equations:

$$\ddot{x}(t) + (M(t) + a_0(t)I_P)\dot{x}(t) + (N(t) + b_0(t)I_p)x(t) = q(t) + c_0(t)e_1, \tag{5.7}$$

where $x(t) = (x_1(t), \ldots, x_P(t))^\top$, I_P is the $P \times P$ identity matrix and e_1 is the first vector of the canonical basis: $(1, 0, \ldots, 0)^\top$. It remains to find the initial condition for (5.7). From $\sum_{i=1}^{P} x_i(t_0)\phi_i(\zeta) = Y_0$ and $\sum_{i=1}^{P} \dot{x}_i(t_0)\phi_i(\zeta) = Y_1$, we obtain that $x_k(t_0) = \mathbb{E}[Y_0\phi_k(\zeta)]$ and $\dot{x}_k(t_0) = \mathbb{E}[Y_1\phi_k(\zeta)]$, for $k = 1, \ldots, P$. Thus, the initial conditions become $x(t_0) = y$ and $\dot{x}(t_0) = y'$, where $y = (y_1, \ldots, y_P)^\top$ and $y' = (y'_1, \ldots, y'_P)^\top$,

$$y_k = \mathbb{E}[Y_0\phi_k(\zeta)], \quad y'_k = \mathbb{E}[Y_1\phi_k(\zeta)], \tag{5.8}$$

for $k = 1, \ldots, P$.

The system of deterministic differential equations can be solved by using standard numerical techniques. Once we have the solution $(x_1(t), \ldots, x_P(t))$, we have obtained the approximation $\sum_{i=1}^{P} x_i(t)\phi_i(\zeta)$ for the solution stochastic process $X(t)$. Moreover, one can approximate the expectation and covariance of $X(t)$:

$$\mathbb{E}[X(t)] \approx x_1(t), \quad \mathbb{C}\mathrm{ov}[X(t_1), X(t_2)] \approx \sum_{i=2}^{P} x_i(t_1)x_i(t_2). \tag{5.9}$$

5.2.2 The random inputs may not be independent

In the notation from [49] and Section 5.1, let $\mathcal{C}_i^p = \{1, \zeta_i, \ldots, \zeta_i^p\}$ be the canonical basis of $\mathcal{P}_p[\zeta_i]$, for $i = 1, \ldots, s$. We construct the basis $\Xi = \{\phi_1, \ldots, \phi_P\}$ of $\mathcal{P}_P^s[\zeta]$ as in [49]. This basis is not orthonormal with respect to the law $\mathbb{P}_\zeta$.

We approximate the solution stochastic process $X(t) \approx \sum_{i=1}^{P} x_i(t)\phi_i(\zeta)$ by imposing the right-hand side to be a solution to the random initial value problem (5.1). One obtains (5.3). By multiplying by $\phi_k(\zeta)$ and applying expectations, $k = 1, \ldots, P$, we derive that

$$\begin{aligned}
&\sum_{i=1}^{P} \ddot{x}_i(t)\mathbb{E}[\phi_i(\zeta)\phi_k(\zeta)] + a_0(t)\sum_{i=1}^{P} \dot{x}_i(t)\mathbb{E}[\phi_i(\zeta)\phi_k(\zeta)] \\
&+ \sum_{i=1}^{d_A}\sum_{j=1}^{P} a_i(t)\dot{x}_j(t)\mathbb{E}[\gamma_i\phi_j(\zeta)\phi_k(\zeta)] + b_0(t)\sum_{i=1}^{P} x_i(t)\mathbb{E}[\phi_i(\zeta)\phi_k(\zeta)] \\
&+ \sum_{i=1}^{d_B}\sum_{j=1}^{P} b_i(t)x_j(t)\mathbb{E}[\eta_i\phi_j(\zeta)\phi_k(\zeta)] \\
&= c_0(t)\mathbb{E}[\phi_k(\zeta)] + \sum_{i=1}^{d_C} c_i(t)\mathbb{E}[\xi_i\phi_k(\zeta)].
\end{aligned} \tag{5.10}$$

Define the $P \times P$ matrix R and the vector h of length P as

$$R_{ik} = \mathbb{E}[\phi_i(\zeta)\phi_k(\zeta)], \quad h_k = \mathbb{E}[\phi_k(\zeta)], \tag{5.11}$$

for $i, k = 1, \ldots, P$. Expression (5.10) can be written in matrix form as a deterministic system of P differential equations:

$$R\ddot{x}(t) + (M(t) + a_0(t)R)\dot{x}(t) + (N(t) + b_0(t)R)x(t) = q(t) + c_0(t)h. \tag{5.12}$$

The initial conditions are given by $Rx(t_0) = y$ and $R\dot{x}(t_0) = y'$.

This system of deterministic differential equations is solvable by standard numerical techniques. Once we have computed the approximation $\sum_{i=1}^{P} x_i(t)\phi_i(\zeta)$ of the solution stochastic process $X(t)$, the expectation and covariance of $X(t)$ can be approximated as follows:

$$\begin{aligned} \mathbb{E}[X(t)] &\approx \sum_{i=1}^{P} x_i(t)\mathbb{E}[\phi_i(\zeta)], \\ \mathbb{C}\mathrm{ov}[X(t_1), X(t_2)] &\approx \sum_{i=1}^{P}\sum_{j=1}^{P} x_i(t_1)x_j(t_2)\mathbb{C}\mathrm{ov}[\phi_i(\zeta), \phi_j(\zeta)]. \end{aligned} \tag{5.13}$$

5.3 Algorithm

In this section we present the algorithm corresponding to Section 5.2. From the random inputs $A(t)$, $B(t)$ and $C(t)$ having expression (5.2) and the initial conditions Y_0 and Y_1, we will show the steps to be followed in order to approximate the expectation and covariance of the solution stochastic process $X(t)$. As in Section 5.2, denote the random input parameters by $\zeta_1, \ldots, \zeta_s$.

Case $\zeta_1, \ldots, \zeta_s$ are independent:

Step 1. Define the canonical basis $\mathcal{C}_i^p = \{1, \zeta_i, \ldots, \zeta_i^p\}$, $i = 1, \ldots, s$.

Step 2. Via a Gram-Schmidt procedure, orthonormalize $\mathcal{C}_i^p$ to a new basis $\Xi_i^p = \{\phi_0^i(\zeta), \ldots, \phi_p^i(\zeta)\}$ with respect to the probability law $\mathbb{P}_{\zeta_i}$ of ζ_i. In the software Mathematica®, this can be readily done with the built-in function `Orthogonalize`. For example, if `p=3` and the probability distribution is `dist`, then the command could be:

```
Expand[Orthogonalize[{1, Z, Z^2, Z^3},
Integrate[#1 #2 PDF[dist, Z], {Z, -Infinity, Infinity}] &]]
```

Step 3. By using a simple tensor product, define the orthonormal basis with respect to the joint law $\mathbb{P}_\zeta = \mathbb{P}_{\zeta_1} \times \cdots \times \mathbb{P}_{\zeta_s}$, $\Xi = \{\phi_1(\zeta), \ldots, \phi_P(\zeta)\}$.

Step 4. Construct the matrices $M(t)$ and $N(t)$ given by (5.5), the vector $q(t)$ defined by (5.6), and the initial conditions y and y' given by (5.8). All the involved expectations can be calculated with the built-in function `Expectation` from Mathematica®.

Step 5. Solve numerically the deterministic system of P differential equations given by (5.7) with initial conditions $x(t_0) = y$ and $\dot{x}(t_0) = y'$. This system does not pose serious numerical challenges. We integrate the equations over time with the standard `NDSolve` routine in Mathematica®: write the instruction

```
NDSolve[eqns,function,{t,t0,T}]
```

with automatic method, step size, etc. (the built-in function will automatically try to estimate the best method for a particular computation).

Step 6. Approximate the expectation and covariance of the unknown solution stochastic process by using (5.9).

Case $\zeta_1, \ldots, \zeta_s$ are not independent:

Step 1. Define the canonical basis $\mathcal{C}_i^p = \{1, \zeta_i, \ldots, \zeta_i^p\}$, $i = 1, \ldots, s$.

Step 2. By using a simple tensor product, define the basis of canonical polynomials $\Xi = \{\phi_1(\zeta), \ldots, \phi_P(\zeta)\}$.

Step 3. Construct the matrices $M(t)$ and $N(t)$ given by (5.5), the vector $q(t)$ defined by (5.6), the matrix $R(t)$ and the vector h given by (5.11), and the vectors y and y' expressed by (5.8). All the involved expectations can be calculated with the built-in function `Expectation` from Mathematica®.

Step 4. Solve numerically the deterministic system of P differential equations given by (5.12) with initial conditions $Rx(t_0) = y$ and $R\dot{x}(t_0) = y'$. This system does not pose serious numerical challenges. We thus integrate the equations over time with the standard `NDSolve` routine from Mathematica® with the option

```
Method -> {"EquationSimplification" -> "Residual"}
```

(to deal with the corresponding system of differential-algebraic equations): write the instruction

```
NDSolve[eqns,function,{t,t0,T},
Method -> {"EquationSimplification" -> "Residual"}]
```

with automatic method, step size, etc. (the built-in function will automatically try to pick the best method for a particular computation).

Step 5. Approximate the expectation and covariance of the unknown solution stochastic process by using (5.13).

5.4 Examples

In this section we show particular examples of the random initial value problem (5.1) to which we apply adaptive gPC to approximate the expectation and covariance of the solution stochastic process $X(t)$.

We will compare the results with Monte Carlo simulation. Sample from the probability distributions of $A(t)$, $B(t)$, $C(t)$, Y_0 and Y_1 to obtain, say m realizations, for m large:

$$A^{(1)}(t),\dots,A^{(m)}(t),\ B^{(1)}(t),\dots,B^{(m)}(t),\ C^{(1)}(t),\dots,C^{(m)}(t),$$

$$Y_0^{(1)},\dots,Y_0^{(m)},\ Y_1^{(1)},\dots,Y_1^{(m)}.$$

Then we solve the m deterministic initial value problems

$$\begin{cases} \ddot{X}^{(i)}(t) + A^{(i)}(t)\dot{X}^{(i)}(t) + B^{(i)}(t)X^{(i)}(t) = C^{(i)}(t),\ t \in \mathbb{R}, \\ X^{(i)}(t_0) = Y_0^{(i)}, \\ \dot{X}^{(i)}(t_0) = Y_1^{(i)}, \end{cases}$$

so that we obtain m realizations of $X(t)$: $X^{(1)}(t),\dots,X^{(m)}(t)$. The law of large numbers permits approximating $\mathbb{E}[X(t)]$ and $\mathbb{V}[X(t)]$ by computing the sample mean and sample variance of $X^{(1)}(t),\dots,X^{(m)}(t)$:

$$\mathbb{E}[X(t)] \approx \mu_m(t) = \frac{1}{m}\sum_{i=1}^{m} X^{(i)}(t),\ \ \mathbb{V}[X(t)] \approx \frac{1}{m-1}\sum_{i=1}^{m}(X^{(i)}(t) - \mu_m(t))^2.$$

The results of adaptive gPC agree with Monte Carlo simulation, although the convergence rate of Monte Carlo is much slower (its error convergence rate is inversely proportional to the square root of the number of realizations [172, p. 53]).

The result of the expectation will also be compared with the dishonest method [85, p. 149]. It consists in estimating $\mathbb{E}[X(t)]$ by substituting $A(t)$, $B(t)$, $C(t)$, Y_0 and Y_1 in (5.1) by their corresponding expected values. Denoting $\mu_X(t) = \mathbb{E}[X(t)]$, the idea is that, since $\mathbb{E}[\ddot{X}(t)] = \frac{\mathrm{d}^2}{\mathrm{d}t^2}(\mu_X(t))$ and $\mathbb{E}[\dot{X}(t)] = \frac{\mathrm{d}}{\mathrm{d}t}(\mu_X(t))$, because of the commutation between the mean square limit and the

expectation operator (see [160, Ch. 4]), one solves:

$$\begin{cases} \frac{\mathrm{d}^2}{\mathrm{d}t^2}(\mu_X(t)) + \mathbb{E}[A(t)]\frac{\mathrm{d}}{\mathrm{d}t}(\mu_X(t)) + \mathbb{E}[B(t)]\mu_X(t) = \mathbb{E}[C(t)],\ t \in \mathbb{R}, \\ \mu_X(t_0) = \mathbb{E}[Y_0], \\ \frac{\mathrm{d}}{\mathrm{d}t}(\mu_X(t_0)) = \mathbb{E}[Y_1]. \end{cases}$$

In our context, the dishonest method will work on cases where $\mathbb{C}\mathrm{ov}[A(t), \dot{X}(t)]$ and $\mathbb{C}\mathrm{ov}[B(t), X(t)]$ are small, but in general, there is no certainty that this may hold. Thus, this method is a naive approximation to the true expectation, with no theoretical support, although with a certain use in the literature [85].

When possible, the results obtained via adaptive gPC for the expectation and variance will be compared with the random Fröbenius method. The convergence of the random Fröbenius method will be guaranteed by Chapter 3.

Several conclusions are drawn from these examples. Adaptive gPC allows for random inputs (5.2) more general than the random Fröbenius method: $A(t)$, $B(t)$ and $C(t)$ may not be analytic, they may be represented via a truncated Karhunen-Loève expansion, etc. Moreover, with a small length p of the bases, accurate results are obtained (this is due to the well-known spectral convergence of gPC-based methods). However, from a computational standpoint, a large number s of random input parameters may make the computations inviable, as the order P of the basis increases as $P = (p+s)!/(p!s!)$.

Example 5.3 Airy's random differential equation is given by [43]:

$$\begin{cases} \ddot{X}(t) + AtX(t) = 0,\ t \in \mathbb{R}, \\ X(0) = Y_0, \\ \dot{X}(0) = Y_1, \end{cases} \tag{5.14}$$

where A, Y_0 and Y_1 are random variables. It is well-known that the solution to the deterministic Airy's differential equation is highly oscillatory, hence it is expected that, in dealing with its stochastic counterpart, differences between distinct methods will be highlighted.

The existence and uniqueness of sample-path solution is guaranteed by Proposition 5.1. Concerning the existence and uniqueness of mean square solution, we refer to Chapter 4.

In [43], the following distributions for A, Y_0 and Y_1 are set: $A \sim \mathrm{Beta}(2,3)$, $Y_0 \sim \mathrm{Normal}(1,1)$ and $Y_1 \sim \mathrm{Normal}(2,1)$. They are assumed to be independent. Approximations for the expectation and variance via the random

Fröbenius method and Monte Carlo simulation are obtained in [43]. We use adaptive gPC (independent case) with $p = 3$, $p = 4$, $\zeta_1 = A$, $\zeta_2 = Y_0$ and $\zeta_3 = Y_1$, $\eta_1 = A$, $A(t) = 0$, $C(t) = 0$ and $b_1(t) = t$. The results obtained are shown in Table 5.1 (expectation), Table 5.2 (variance) and Table 5.3 (covariance). The order of truncation in the random Fröbenius method is denoted by N. Observe that gPC expansions have converged for $t \in [0, 2]$ with order $p = 3$. This rapid convergence shows the potentiality of this approach.

In Figure 5.1, we focus on the convergence of gPC expansions. The solid line reflects the expectations, while the dashed lines represent confidence intervals constructed with the rule mean $\pm$ deviation (the standard deviation stands for the square root of the variance). Observe that, as we move away from $t = 0$, larger orders of p are required to achieve good approximations of the statistics of $X(t)$. Indeed, Galerkin projections deviate from the exact solution after a certain time. Realize also that larger orders of p are needed to get accurate results of the standard deviation than for the expectation (statistical moments of order 2 are harder to approximate than moments of order 1). For $p = 3$ and $p = 4$, the approximate expectations agree up to time $t = 7$, whereas the standard deviations up to $t = 4.5$. For $p = 2$ and $p = 3$, similar means are obtained until $t = 6$, and similar standard deviations up to $t = 4$. Notice that the convergence deteriorates for $p = 1$: the results for $p = 1$ and $p = 2$ agree until $t = 4$ for the expectation, but up to instant $t = 1.5$ for the standard deviation. As p grows, the approximation of the statistics will improve for larger t.

t	gPC $p = 3$	gPC $p = 4$	Fröb. $N = 3$	Fröb. $N = 5$	dishonest	MC $50,000$	MC $100,000$
0.00	1		1	1	1	0.99701	1.00138
0.25	1.49870	1.49870	1.49870	1.49870	1.49870	1.49519	1.49976
0.50	1.98752	1.98752	1.98752	1.98752	1.98752	1.98353	1.98829
0.75	2.45108	2.45108	2.45108	2.45108	2.45102	2.44667	2.45160
1.00	2.86856	2.86856	2.86856	2.86856	2.86818	2.86383	2.86893
1.25	3.21494	3.21494	3.21494	3.21494	3.21339	3.21008	3.21534
1.50	3.46310	3.46310	3.46310	3.46310	3.45812	3.45831	3.46376
1.75	3.58660	3.58660	3.58660	3.58660	3.57340	3.58215	3.58784
2.00	3.56335	3.56335	3.56336	3.56335	3.53286	3.55948	3.56552

Table 5.1: Approximation of $\mathbb{E}[X(t)]$. Example 5.3, assuming independent random data.

By using the random Fröbenius method from [43], an example of Airy's differential equation with dependent random inputs is performed. It is set (A, Y_0, Y_1) to have a multivariate Gaussian distribution, with mean vector and covariance

t	gPC $p=3$	gPC $p=4$	Fröb. $N=3$	Fröb. $N=5$	MC $50,000$	MC $100,000$
0.00	1	1	1	1	0.99610	0.99530
0.25	1.06035	1.06035	1.06035	1.06035	1.05902	1.05642
0.50	1.23142	1.23142	1.23142	1.23142	1.23408	1.22793
0.75	1.49261	1.49261	1.49261	1.49261	1.50041	1.48944
1.00	1.81392	1.81392	1.81392	1.81392	1.82744	1.81127
1.25	2.15870	2.15870	2.15870	2.15870	2.17768	2.15721
1.50	2.49379	2.49379	2.49379	2.49379	2.51690	2.49462
1.75	2.80560	2.80560	2.80560	2.80560	2.83029	2.81030
2.00	3.11530	3.11530	3.11530	3.11530	3.13783	3.12559

Table 5.2: Approximation of $\mathbb{V}[X(t)]$. Example 5.3, assuming independent random data.

t–s	0	0.25	0.5	0.75	1	1.25	1.5	1.75	2
0	1.	0.998959	0.991684	0.972072	0.934436	0.873965	0.787323	0.673299	0.533429
0.25	0.998959	1.06035	1.11507	1.15586	1.17516	1.16565	1.12103	1.03694	0.911972
0.5	0.991684	1.11507	1.23142	1.33242	1.40874	1.45067	1.44906	1.39647	1.28856
0.75	0.972072	1.15586	1.33242	1.49261	1.62561	1.7196	1.76286	1.74509	1.65909
1	0.934436	1.17516	1.40874	1.62561	1.81392	1.96032	2.05099	2.07321	2.01713
1.25	0.873965	1.16565	1.45067	1.7196	1.96032	2.1587	2.2997	2.3688	2.35387
1.5	0.787323	1.12103	1.44906	1.76286	2.05099	2.2997	2.49379	2.61793	2.65822
1.75	0.673299	1.03694	1.39647	1.74509	2.07321	2.3688	2.61793	2.8056	2.91699
2	0.533429	0.911972	1.28856	1.65909	2.01713	2.35387	2.65822	2.91699	3.1153

Table 5.3: Approximation of $\mathbb{C}\mathrm{ov}[X(t), X(s)]$ via adapted gPC with $p=3$ and $p=4$. Example 5.3, assuming independent random data.

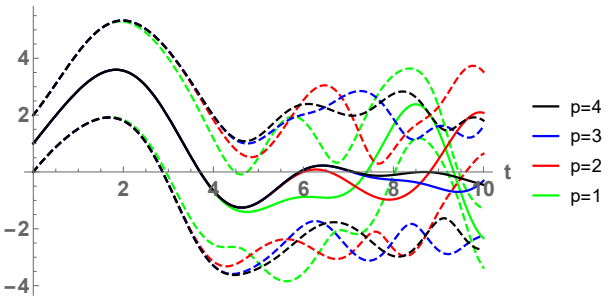

Figure 5.1: Expectation and confidence interval for the solution stochastic process, for orders of basis $p=1,2,3,4$. Example 5.3, assuming independent random data.

matrix given by

$$\mu = \begin{pmatrix} 0.4 \\ 1 \\ 2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 0.04 & 0.0001 & -0.05 \\ 0.0001 & 1 & 0.5 \\ -0.005 & 0.5 & 1 \end{pmatrix},$$

respectively. In Table 5.4, Table 5.5 and Table 5.6, the results obtained via adaptive gPC with $p = 3$, $p = 4$ (dependent case) and [43] are shown. Adaptive gPC converges for small order of basis p.

In Figure 5.2, we analyze the convergence of gPC expansions by depicting the expectation (solid line) and confidence interval (dashed lines) for $X(t)$, where the confidence interval is constructed as mean $\pm$ deviation. Analogous comments to those from Figure 5.1 apply in this case again. For orders $p = 3$ and $p = 4$, the expectations agree up to time $t = 6$, while the standard deviations coincide until $t = 4.6$. For $p = 2$ and $p = 3$, the means are similar until $t = 6$, whereas the dispersion estimates start separating from $t = 3.8$. Finally, for $p = 1$ and $p = 2$, the approximations for the average statistic coincide till $t = 4.5$, and for the deviation statistic until $t = 2.5$.

t	gPC $p=3$	gPC $p=4$	Fröb. $N=4$	Fröb. $N=5$	dishonest	MC $50,000$	MC $100,000$
0.00	1	1	1	1	1	1.00188	1.00597
0.25	1.49870	1.49870	1.49870	1.49870	1.49870	1.50166	1.50581
0.50	1.98755	1.98755	1.98755	1.98755	1.98752	1.99156	1.99575
0.75	2.45121	2.45121	2.45121	2.45121	2.45102	2.45622	2.46041
1.00	2.86895	2.86895	2.86895	2.86895	2.86818	2.87485	2.87900
1.25	3.21589	3.21589	3.21589	3.21589	3.21339	3.22247	3.22656
1.50	3.46503	3.46503	3.46503	3.46503	3.45812	3.47198	3.47601
1.75	3.59010	3.59010	3.59010	3.59010	3.57340	3.59700	3.60101
2.00	3.56914	3.56914	3.56915	3.56914	3.53286	3.57546	3.57949

Table 5.4: Approximation of $\mathbb{E}[X(t)]$. Example 5.3, assuming dependent random data.

Example 5.4 Consider the random differential equation

$$\begin{cases} \ddot{X}(t) + (\gamma_1 + \gamma_2 t)\dot{X}(t) + (\eta_1 + t)X(t) = \xi_1 \cos(t) + g(t), \ t \in \mathbb{R}, \\ X(0) = Y_0, \\ \dot{X}(0) = Y_1, \end{cases} \tag{5.15}$$

where $\gamma_1 \sim$ Poisson(3), $\gamma_2 \sim$ Uniform$(0,1)$, $\eta_1 \sim$ Gamma$(2,2)$, $Y_0 = -1$, $Y_1 \sim$ Exponential(4), $\xi_1 \sim$ Uniform$(-8,2)$ and $g(t) = \mathrm{e}^{-1/t} 1_{(0,\infty)}(t)$.

Proposition 5.1 ensures the existence and uniqueness of a sample-path solution. To apply Proposition 5.2, one would need to truncate the supports of γ_1 and

t	gPC $p=3$	gPC $p=4$	Fröb. $N=3$	Fröb. $N=4$	MC $50,000$	MC $100,000$
0.00	1	1	1	1	0.999223	0.99992
0.25	1.30997	1.30997	1.30997	1.30997	1.30713	1.30991
0.50	1.72535	1.72535	1.72535	1.72535	1.71989	1.72525
0.75	2.21241	2.21241	2.21241	2.21241	2.20395	2.21230
1.00	2.72122	2.72122	2.72122	2.72122	2.70957	2.72125
1.25	3.19236	3.19236	3.19236	3.19236	3.17745	3.19283
1.50	3.57361	3.57361	3.57361	3.57361	3.55537	3.57484
1.75	3.84459	3.84459	3.84454	3.84458	3.82262	3.84669
2.00	4.04087	4.04087	4.04090	4.04086	4.01420	4.04342

Table 5.5: Approximation of $\mathbb{V}[X(t)]$. Example 5.3, assuming dependent random data.

t–s	0	0.25	0.5	0.75	1	1.25	1.5	1.75	2
0	1.	1.12389	1.24064	1.34181	1.41793	1.45914	1.45614	1.40144	1.29068
0.25	1.12389	1.30997	1.48771	1.64683	1.77533	1.86031	1.88919	1.85125	1.7394
0.5	1.24064	1.48771	1.72535	1.94152	2.12187	2.25061	2.31202	2.29226	2.18163
0.75	1.34181	1.64683	1.94152	2.21241	2.4431	2.61534	2.71062	2.71235	2.60832
1	1.41793	1.77533	2.12187	2.4431	2.72122	2.93619	3.06742	3.09609	3.00785
1.25	1.45914	1.86031	2.25061	2.61534	2.93619	3.19236	3.36219	3.42552	3.36639
1.5	1.45614	1.88919	2.31202	2.71062	3.06742	3.36219	3.57361	3.68135	3.66862
1.75	1.40144	1.85125	2.29226	2.71235	3.09609	3.42552	3.68135	3.84459	3.89856
2	1.29068	1.7394	2.18163	2.60832	3.00785	3.36639	3.66862	3.89856	4.04087

Table 5.6: Approximation of $\mathbb{C}\mathrm{ov}[X(t), X(s)]$ via adapted gPC with $p=3$ and $p=4$. Example 5.3, assuming dependent random data.

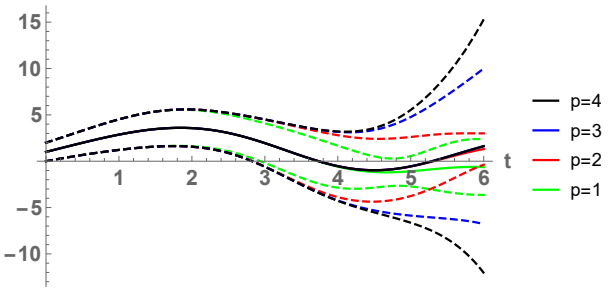

Figure 5.2: Expectation and confidence interval for the solution stochastic process, for orders of basis $p=1,2,3,4$. Example 5.3, assuming dependent random data.

η_1. These truncations can be constructed on intervals as large as desired, in order to maintain the results.

The input random variables $\zeta_1 = \gamma_1$, $\zeta_2 = \gamma_2$, $\zeta_3 = \eta_1$, $\zeta_4 = \xi_1$ and $\zeta_5 = Y_1$ are assumed to be independent. The involved functions are $a_1(t) = 1$, $a_2(t) = t$, $b_1(t) = 1$, $b_2(t) = t$, $c_0(t) = g(t)$ and $c_1(t) = \cos(t)$. Notice that $C(t)$ is not an analytic stochastic process, because $g(t)$ is not a real analytic function. The random Fröbenius method is not applicable for the random initial value problem (5.15). However, we are going to see that adaptive gPC (independent case) with $p = 6$ and $p = 7$ provides reliable approximations of the expectation and covariance of $X(t)$. We will compare the results with Monte Carlo simulation. In Table 5.7, Table 5.8 and Table 5.9 we show the estimates obtained.

In Figure 5.3, we focus on the convergence of gPC expansions. We depict the estimates of the expectations (solid line) and confidence intervals (dashed lines), with the rule mean $\pm$ deviation, for orders $p = 4, 5, 6, 7$. Note that convergence is achieved for $t \in [0, 10]$.

t	gPC $p = 6$	gPC $p = 7$	dishonest	MC $50,000$	MC $100,000$
0.00	-1	-1	-1	-1	-1
0.25	-0.930972	-0.930972	-0.931372	-0.931035	-0.930364
0.50	-0.855779	-0.855779	-0.852372	-0.855937	-0.854386
0.75	-0.780021	-0.780021	-0.759103	-0.780573	-0.778022
1.00	-0.700758	-0.700758	-0.647653	-0.702042	-0.698437
1.25	-0.609156	-0.609156	-0.518169	-0.611266	-0.606832
1.50	-0.496445	-0.496446	-0.374486	-0.499156	-0.494407
1.75	-0.359632	-0.359635	-0.222874	-0.362532	-0.358036
2.00	-0.203726	-0.203737	-0.070806	-0.206408	-0.202560

Table 5.7: Approximation of $\mathbb{E}[X(t)]$. Example 5.4, assuming independent random data.

t	gPC $p = 6$	gPC $p = 7$	MC $50,000$	MC $100,000$
0.00	0	0	0	0
0.25	0.0114271	0.0114271	0.0115378	0.0114953
0.50	0.0897916	0.0897916	0.0904703	0.090160
0.75	0.236135	0.236136	0.237288	0.237066
1.00	0.371625	0.371639	0.372899	0.373058
1.25	0.426921	0.427021	0.428690	0.428342
1.50	0.388485	0.388899	0.391305	0.38978
1.75	0.289622	0.290720	0.293631	0.291429
2.00	0.182954	0.184922	0.187906	0.185917

Table 5.8: Approximation of $\mathbb{V}[X(t)]$. Example 5.4, assuming independent random data.

t–s	0	0.25	0.5	0.75	1	1.25	1.5	1.75	2
0	0	0	0	0	0	0	0	0	0
0.25	0	0.0114271	0.0310997	0.0486065	0.0585142	0.0597207	0.0537473	0.042977	0.0295893
0.5	0	0.0310997	0.0897916	0.14431	0.176903	0.182770	0.165621	0.132554	0.0905903
0.75	0	0.0486065	0.14431	0.236136	0.293929	0.307634	0.281495	0.226547	0.154877
1	0	0.0585142	0.176903	0.293929	0.371639	0.394991	0.366485	0.29839	0.206021
1.25	0	0.0597207	0.182770	0.307634	0.394991	0.427021	0.403260	0.334311	0.235733
1.5	0	0.0537473	0.165621	0.281495	0.366485	0.403260	0.388899	0.330575	0.241178
1.75	0	0.042977	0.132554	0.226547	0.29839	0.334311	0.330575	0.29072	0.223109
2	0	0.0295893	0.0905903	0.154877	0.206021	0.235733	0.241178	0.223109	0.184922

Table 5.9: Approximation of $\mathbb{C}\mathrm{ov}[X(t), X(s)]$ via adapted gPC with $p = 7$. Example 5.4, assuming independent random data.

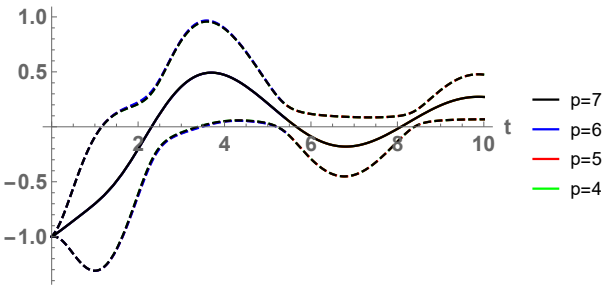

Figure 5.3: Expectation and confidence interval for the solution stochastic process, for orders of basis $p = 4, 5, 6, 7$. Example 5.4, assuming independent random data.

Example 5.5 Consider the random differential equation

$$\begin{cases} \ddot{X}(t) + B(t)X(t) = C,\ t \in [0,1], \\ X(0) = Y_0, \\ \dot{X}(0) = Y_1, \end{cases} \tag{5.16}$$

where $B(t)$ is a standard Brownian motion on $[0,1]$, $C \sim \text{Poisson}(2)$, and the initial conditions are distributed as $Y_0 \sim \text{Beta}(1/2, 1/2)$ and $Y_1 = 0$. These random inputs are assumed to be independent.

This stochastic system has a unique solution in the sample-path sense, by Proposition 5.1. In principle, one cannot ensure the existence of a mean square solution, since the sample paths of Brownian motion are not bounded.

Consider the Karhunen-Loève expansion of Brownian motion [119, p. 216]:

$$B(t) = \sum_{j=1}^{\infty} \frac{\sqrt{2}}{\left(j - \frac{1}{2}\right)\pi} \sin\left(\left(j - \frac{1}{2}\right)\pi t\right) \xi_j,$$

where $\xi_1, \xi_2, \ldots$ are independent and Normal$(0,1)$ random variables. The series is understood in $\mathrm{L}^2([0,1] \times \Omega)$. We truncate the Karhunen-Loève expansion so that $B(t)$ will have the form in (5.2). If we take $d_B = 7$, we are capturing more than 97% of the total variance of X. Thus, we take

$$B(t) = \sum_{j=1}^{7} \frac{\sqrt{2}}{\left(j - \frac{1}{2}\right)\pi} \sin\left(\left(j - \frac{1}{2}\right)\pi t\right) \xi_j.$$

The random inputs become $\zeta_1 = \xi_1, \ldots, \zeta_7 = \xi_7$, $\zeta_8 = C$ and $\zeta_9 = Y_0$, with functions $b_j(t) = \frac{\sqrt{2}}{(j-1/2)\pi} \sin((j-1/2)\pi t)$, $1 \le j \le 7$, and $c_1(t) = 1$.

Notice that, if one truncates $\xi_1, \ldots, \xi_7$ to a large but bounded support, Proposition 5.2 entails that there exists a solution stochastic process in the mean square sense.

In Table 5.10, Table 5.11 and Table 5.12, we show the results obtained by adaptive gPC with $p = 2$, $p = 3$ and Monte Carlo simulation. Similar estimates are obtained for $p = 2$ and $p = 3$, which agrees with the convergence of gPC-based representations.

In Figure 5.4, we show graphically the convergence of gPC expansions on $[0,1]$: we plot the approximate expectations (solid line) and confidence intervals (dashed lines) for $X(t)$, where the confidence interval is constructed as mean $\pm$ deviation. For $p = 1, 2, 3$, no differences in the estimates are observed.

t	gPC $p=2$	gPC $p=3$	dishonest	MC $50,000$	MC $100,000$
0.00	0.5	0.5	0.5	0.499302	0.499056
0.25	0.562504	0.562504	0.5625	0.561732	0.561438
0.50	0.75014	0.75014	0.75	0.749196	0.748740
0.75	1.06365	1.06365	1.0625	1.06245	1.06179
1.00	1.50536	1.50536	1.5	1.50396	1.50311

Table 5.10: Approximation of $\mathbb{E}[X(t)]$. Example 5.5, assuming independent random data.

t	gPC $p=2$	gPC $p=3$	MC $50,000$	MC $100,000$
0.00	0.125	0.125	0.124849	0.124826
0.25	0.126974	0.126974	0.126883	0.126841
0.50	0.157008	0.157008	0.157267	0.157033
0.75	0.290263	0.290265	0.291811	0.290766
1.00	0.664551	0.664592	0.670067	0.666466

Table 5.11: Approximation of $\mathbb{V}[X(t)]$. Example 5.5, assuming independent random data.

t–s	0	0.25	0.5	0.75	1
0	0.125	0.125001	0.125033	0.125248	0.126043
0.25	0.125001	0.126974	0.132949	0.143104	0.157885
0.5	0.125033	0.132949	0.157008	0.197634	0.255578
0.75	0.125248	0.143104	0.197634	0.290265	0.422829
1	0.126043	0.157885	0.255578	0.422829	0.664592

Table 5.12: Approximation of $\mathbb{C}\text{ov}[X(t), X(s)]$ via adapted gPC with $p=3$. Example 5.5, assuming independent random data.

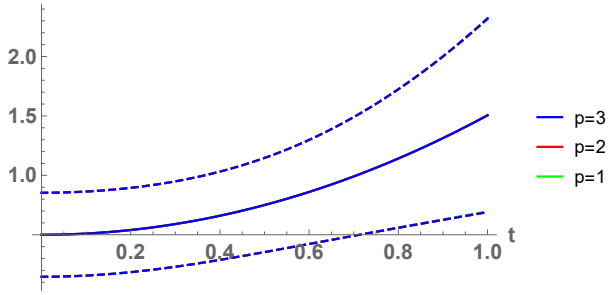

Figure 5.4: Expectation and confidence interval for the solution stochastic process, for orders of basis $p=1,2,3$. Example 5.5, assuming independent random data.

5.5 Conclusions

In this chapter, we have quantified computationally the uncertainty of random non-autonomous second-order linear differential equations via adaptive gPC. After reviewing adaptive gPC from the extant literature, we have provided a methodology and an algorithm to approximate computationally the expectation and covariance of the solution stochastic process. The hypotheses from our algorithm allow both independent and dependent random parameter inputs, being both absolutely continuous or discrete with infinitely many point masses. The generality of our computational results allows the random input coefficients to be truncated random power series or truncated Karhunen-Loève expansions. The former case permits comparing our methodology with the random Fröbenius method, an approach already used in the literature with particular random second-order linear differential equations, and Monte Carlo simulation. A wide variety of examples show that adaptive gPC becomes successful quantifying the uncertainty of random non-autonomous second-order linear differential equations, even when the random Fröbenius method is not applicable.

Acknowledgements

This work has been supported by the Spanish Ministerio de Economía y Competitividad grant MTM2017–89664–P. I acknowledge the doctorate scholarship granted by Programa de Ayudas de Investigación y Desarrollo (PAID), Universitat Politècnica de València.

The main results of this chapter have been published in [22].

Chapter 6

Analysis of second-order linear differential equations with analytic uncertainties via the computation of the density function

Usually, solving random second-order linear differential equations consists of computing the first statistics of the response process, and that task has been an essential goal in the literature. A more ambitious objective is the computation of the solution probability density function. We present advances on these two aspects in the general case of analytic data processes. The Fröbenius method is employed to obtain the stochastic solution in the form of a mean square convergent power series. We rely on the law of total probability to express the density function in closed-form as an expectation. For the computation of this expectation, a sequence of approximating density functions is constructed by reducing the dimensionality of the problem using the truncated power series of the fundamental set. We prove several theoretical results regarding the pointwise convergence of the sequence of density functions and the convergence in total variation. The pointwise convergence turns out to be exponential under a Lipschitz hypothesis. As the density functions are expressed in terms of expectations, we propose a symbolic Monte Carlo sampling algorithm for their estimation. This algorithm is applied on several examples.

6.1 Introduction

For random second-order linear differential equations, important advances have been achieved for the computation of the first moments of the solution, via mean square calculus and the so-called Fröbenius method. The general stochastic system is given by (3.1).

A more ambitious objective is the computation of the probability density function of $X(t)$, denoted hereafter as $f_{X(t)}(x) = \frac{\mathrm{d}(\mathbb{P}\circ X(t)^{-1})(x)}{\mathrm{d}x}$. In [38], the authors constructed approximations of the probability density functions of the solution to (3.1) when $A(t)$ and $B(t)$ do not vary stochastically in time, that is, when $A(t) = A$ and $B(t) = B$ are actually absolutely continuous random variables (autonomous case). A recent paper, [48], presents the approximation of the probability density function of $X(t)$ when $A(t) = p(t; D)$ and $B(t) = q(t; D)$, that is to say, when both $A(t)$ and $B(t)$ depend on a unique absolutely continuous random variable D. This approach does not extend to the general problem (3.1) and certain theoretical points from that contribution are unclear.

In this work, we provide an analysis of (3.1) via the Fröbenius method. The solution is expressed in the form of a mean square convergent power series, under $\mathrm{L}^\infty(\Omega)$ convergence of $A(t) = \sum_{n=0}^{\infty} A_n(t-t_0)^n$ and $B(t) = \sum_{n=0}^{\infty} B_n(t-t_0)^n$ and mean square integrability of the initial data Y_0 and Y_1. Truncation of unbounded supports of random coefficients can be carried out to assure the required boundedness. The bias error of the Fröbenius method is proved to decrease exponentially with the number of terms in the series. Therefore rapid approximations of the statistical moments of $X(t)$ can be derived. However, the exponential convergence is not uniform in time, and it may deteriorate as we move away from the initial instant t_0. Section 6.2 considers these issues. An additional issue is the computation of the probability density function of $X(t)$. Theoretically, it is given by a closed-form expression in terms of an expectation derived from the law of total probability and by exploiting the linearity of the problem. However, to evaluate it in practice, a dimensionality reduction of the problem is required. By truncating the power series, we construct a sequence of probability density functions that, under certain assumptions regarding Nemytskii operators, converges to the target density function pointwise. In this setting, the pointwise convergence of the densities implies convergence in $\mathrm{L}^1(\mathbb{R})$ (total variation distance), and in fact, in $\mathrm{L}^p(\mathbb{R})$, for $1 \leq p < \infty$. The pointwise convergence rate is proved to be exponential under a certain Lipschitz condition, albeit being again not uniform in time. This theoretical analysis on the approximation of the probability density function is

presented in Section 6.3. As each approximating density function is expressed in terms of an expectation, they can be estimated via a Monte Carlo sampling strategy. The procedure is implemented in the form of a symbolic algorithm, whose computational aspects are detailed in Section 6.4. The proposed algorithm is tested on several numerical examples in Section 6.5, to verify the theoretical findings of the chapter and to illustrate computational aspects. Finally, Section 6.6 draws the main conclusions and points out potential lines of research for the future.

6.2 Stochastic solution

The initial value problem (3.1) was studied in the mean square sense in Chapter 3. The theory can be extended to a more general convergence measure, by considering $\mathrm{L}^p(\Omega)$ convergence, $1 \leq p \leq \infty$: if $A(t)$ and $B(t)$ are two random power series with convergence in $\mathrm{L}^\infty(\Omega)$, for $t \in (t_0 - r, t_0 + r)$, and the initial conditions Y_0 and Y_1 belong to $\mathrm{L}^p(\Omega)$, then the stochastic process $X(t) = \sum_{n=0}^{\infty} X_n (t - t_0)^n$ is the $\mathrm{L}^p(\Omega)$ solution to (3.1) on $(t_0 - r, t_0 + r)$.

Regarding the rapidity of convergence of the power series $X(t) = \sum_{n=0}^{\infty} X_n (t - t_0)^n$ introduced in Theorems 3.3 and 3.4, some theoretical estimates were obtained in Chapter 3, although no rate of convergence was derived. Fixed $r > 0$ finite, given $\rho := |t - t_0| < r$ and given an arbitrary s such that $\rho < s < r$, the following estimate holds:

$$\|X^N(t) - X(t)\|_2 \leq K\left(r, s, \{\|A_i\|_\infty\}_{i=1}^{\infty}, \{\|B_i\|_\infty\}_{i=1}^{\infty}, \|Y_0\|_2, \|Y_1\|_2\right) \cdot \frac{(\rho/s)^{N+1}}{1 - \rho/s}.$$

In particular, estimates of $|\mathbb{E}[X^N(t)] - \mathbb{E}[X(t)]|$ and $|\mathbb{V}[X^N(t)] - \mathbb{V}[X(t)]|$ in terms of $\|X^N(t) - X(t)\|_2$ are derived by using Jensen's and Cauchy-Schwarz inequalities. In general, it holds for $1 \leq p \leq \infty$,

$$\|X^N(t) - X(t)\|_p \leq K\left(r, s, \{\|A_i\|_\infty\}_{i=1}^{\infty}, \{\|B_i\|_\infty\}_{i=1}^{\infty}, \|Y_0\|_p, \|Y_1\|_p\right) \cdot \frac{(\rho/s)^{N+1}}{1 - \rho/s}.$$

The constant K can be constructed as follows (see Chapter 3):

Step 1. Given $u = (r + s)/2 \in (s, r)$, choose a constant $C_u > 0$ such that $\|A_i\|_\infty \leq C_u/u^i$ and $\|B_i\|_\infty \leq C_u/u^i$, $i \geq 0$. Such a constant C_u exists because $\sum_{i=0}^{\infty} \|A_i\|_\infty u^i < \infty$ and $\sum_{i=0}^{\infty} \|B_i\|_\infty u^i < \infty$.

Step 2. Pick an integer $n \geq 0$ such that

$$\frac{ns}{(n+2)u} + \frac{C_u s}{n+2} + \frac{C_u s^2}{(n+2)(n+1)} < 1.$$

Step 3. Take $K = \max_{0\leq m\leq n} H_m s^m$, where $\{H_m\}_{m=0}^{\infty}$ satisfies the following recursive equation:

$$H_{m+2} = \left(\frac{m}{(m+2)u} + \frac{C_u}{m+2}\right) H_{m+1} + \frac{C_u}{(m+2)(m+1)} H_m,\ m \geq 0,$$
$$H_0 = \|Y_0\|_2,\ H_1 = \|Y_1\|_2.$$

From the constructed K and given a target error $\epsilon > 0$, a truncation order

$$N > \frac{\log\left(\epsilon^{-1}K(1-\rho/s)^{-1}\right)}{\log(s/\rho)} - 1 = \mathcal{O}\left(\log(\epsilon^{-1})\right)$$

guarantees a root mean square error $\|X^N(t) - X(t)\|_2$ less than ϵ. The number s is arbitrary in (ρ, r). Unfortunately, we are not aware of any method to choose the optimal $s \in (\rho, r)$ minimizing $N = \log(\epsilon^{-1}K(1-\rho/s)^{-1})/\log(s/\rho)$.

We stress several new consequences from these estimates. First, the rate of convergence of $\{X^N(t)\}_{N=0}^{\infty}$ towards $X(t)$ as $N \to \infty$ is exponential, for $t \in (t_0 - r, t_0 + r)$, because it is proportional to $(\rho/s)^N$. Second, the convergence rate may deteriorate severely for large $\rho = |t - t_0|$ and large norms of the random input coefficients. Indeed, K is growing with $s > \rho$ and the norms.

The fact that the convergence rate deteriorates for large $|t-t_0|$ is clear. Assume that $\sum_{n=N}^{\infty} \|X_n\|_2 |t-t_0|^n < \epsilon$, for some target error $\epsilon > 0$. As $|t-t_0|^n$ increases when $|t - t_0|$ grows, a larger N is needed to achieve a root mean square error less than ϵ. This fact may especially occur for $|t - t_0| \geq 1$, as in this case $|t - t_0|^n$ does not tend to 0 when $n \to \infty$, therefore a faster decay of the coefficients $\|X_n\|_2$ is needed to assure convergence.

The numerical experiments presented in Chapter 3 also permitted analyzing the behavior of convergence. As theoretically expected by our exposition, the most important phenomenon observed was that the convergence rate deteriorates severely when the distance $|t - t_0|$ increases, therefore making the Fröbenius method computationally intractable. This issue also occurs with PC-type methods, which require large orders for long-time integration [74]. Nonetheless, for not too large $|t - t_0|$ (this "not too large" is problem-dependent), the Fröbenius method works very well.

6.3 Computation of the probability density function

We now turn to the computation of the probability density function of $X(t)$. Having clarified the conditions for the existence of the solution in Chapter 3, we start by rewriting $X(t)$ in an alternative form.

Theorem 6.1 *Let $A(t) = \sum_{n=0}^{\infty} A_n (t-t_0)^n$ and $B(t) = \sum_{n=0}^{\infty} B_n (t-t_0)^n$ be two random series in the $\mathrm{L}^\infty(\Omega)$ setting, for $t \in (t_0 - r, t_0 + r)$, being $r > 0$ finite and fixed. Assume that the initial conditions Y_0 and Y_1 belong to $\mathrm{L}^2(\Omega)$. Then the mean square analytic solution $X(t)$ can be expressed as $X(t) = Y_0 S_0(t) + Y_1 S_1(t)$, $t \in (t_0 - r, t_0 + r)$, where $S_0(t)$ and $S_1(t)$ are random power series solutions in $\mathrm{L}^\infty(\Omega)$ to*

$$\begin{cases} \ddot{S}_0(t) + A(t)\dot{S}_0(t) + B(t)S_0(t) = 0, \ t \in (t_0 - r, t_0 + r), \\ S_0(t_0) = 1, \\ \dot{S}_0(t_0) = 0, \end{cases}$$

$$\begin{cases} \ddot{S}_1(t) + A(t)\dot{S}_1(t) + B(t)S_1(t) = 0, \ t \in (t_0 - r, t_0 + r), \\ S_1(t_0) = 0, \\ \dot{S}_1(t_0) = 1. \end{cases}$$

Notice that we have written $X(t)$ as a linear combination of the fundamental set $\{S_0(t), S_1(t)\}$. This expression exploits the linearity of the problem. The processes $S_0(t)$ and $S_1(t)$ are random power series in $\mathrm{L}^\infty(\Omega)$,

$$S_0(t) = \sum_{n=0}^{\infty} S_{0,n}(t-t_0)^n, \quad S_1(t) = \sum_{n=0}^{\infty} S_{1,n}(t-t_0)^n,$$

whose coefficients satisfy a difference equation as in Theorem 3.3:

$$S_{0,0} = 1, \quad S_{0,1} = 0,$$

$$S_{0,n+2} = \frac{-1}{(n+2)(n+1)} \sum_{m=0}^{n} \left[(m+1) A_{n-m} S_{0,m+1} + B_{n-m} S_{0,m} \right], \ n \geq 0,$$

and

$$S_{1,0} = 0, \quad S_{1,1} = 1,$$

$$S_{1,n+2} = \frac{-1}{(n+2)(n+1)} \sum_{m=0}^{n} \left[(m+1) A_{n-m} S_{1,m+1} + B_{n-m} S_{1,m} \right], \ n \geq 0.$$

These coefficients can be easily evaluated numerically.

The following lemma is necessary to compute the probability density function $f_{X(t)}(x)$. Its proof is a consequence of the law of total probability [4, Ch. 6], [128, Def. 7.11].

Lemma 6.2 *Let U be an absolutely continuous random variable, independent of the random vector (Z_1, Z_2), where $Z_1 \neq 0$ almost surely. Then $Z_1U + Z_2$ is absolutely continuous, with density function*

$$f_{Z_1U+Z_2}(z) = \mathbb{E}\left[f_U\left(\frac{z - Z_2}{Z_1}\right)\frac{1}{|Z_1|}\right].$$

This lemma provides an alternative to the random variable transformation method [48, Th. 1], in the case of affine mappings. It does not require that the random quantities have an absolutely continuous probability law, a fact that presents advantages from the practical perspective. The drawback is that we need independence between U and (Z_1, Z_2) to represent the probability density function as an expectation. The expectation can be approximated via sampling-based statistical methods, as discussed later on.

The following theorem, which derives the probability density function of $X(t)$, is a straightforward consequence of Lemma 6.2.

Theorem 6.3 *Let $A(t) = \sum_{n=0}^{\infty} A_n(t - t_0)^n$ and $B(t) = \sum_{n=0}^{\infty} B_n(t - t_0)^n$ be two random series in the $\mathrm{L}^\infty(\Omega)$ setting, for $t \in (t_0 - r, t_0 + r)$, being $r > 0$ finite and fixed. Suppose that the initial conditions Y_0 and Y_1 belong to $\mathrm{L}^2(\Omega)$. If $S_0(t) \neq 0$ almost surely, if Y_0 is absolutely continuous, with density function f_{Y_0}, and it is independent of the rest of random input parameters of (3.1), then the mean square solution $X(t)$ has for probability density function*

$$f_{X(t)}(x) = \mathbb{E}\left[f_{Y_0}\left(\frac{x - Y_1S_1(t)}{S_0(t)}\right)\frac{1}{|S_0(t)|}\right]. \tag{6.1}$$

An important issue of expression (6.1) is that $S_0(t)$ and $S_1(t)$ are given by infinite series, therefore truncated approximations are needed. We have to justify that it is legitimate to reduce dimensionality and use truncated random power series for $S_0(t)$ and $S_1(t)$. In what follows, we denote by $S_0^N(t)$ and $S_1^N(t)$ the N-th partial sums of $S_0(t)$ and $S_1(t)$, respectively, which converge in $\mathrm{L}^\infty(\Omega)$ for each t. Let $X^N(t) = Y_0S_0^N(t) + Y_1S_1^N(t)$ be a truncation of the solution $X(t)$, which converges in the mean square sense for each time t.

In the study of random differential equation problems with no closed-form expression of the solution process but only an infinite expansion, one usually

constructs an approximating sequence of stochastic processes with reduced dimensionality and computable probability density function. One thus obtains an approximating sequence of probability density functions, which hopefully presents rapid convergence to the target density function. Moreover, the discontinuity and non-differentiability points of the target density function are captured with no difficulty. In the literature, one may find applications of this type of strategy with power series and Karhunen-Loève developments [19, 20], finite difference schemes [62, 63], and gPC expansions [14].

If $S_0^N(t) \neq 0$ almost surely, then the probability density function of the truncation $X^N(t)$ is

$$f_{X^N(t)}(x) = \mathbb{E}\left[f_{Y_0}\left(\frac{x - Y_1 S_1^N(t)}{S_0^N(t)}\right)\frac{1}{|S_0^N(t)|}\right]. \tag{6.2}$$

This expression involves a maximum of $2N+3$ random variables (Y_1, $S_{0,n}$ and $S_{1,n}$ for $0 \leq n \leq N$). Thus, the expectation can be computed by numerical integration (in the case of absolutely continuous random input coefficients), or by a Monte Carlo procedure [172, pp. 53–54], by sampling realizations of Y_1, $S_0^N(t)$ and $S_1^N(t)$. This is the same strategy as the one followed in our recent paper [20]. The approach based on numerical integration would be feasible only in the case of small N and $A(t) = p(t;D)$, $B(t) = q(t;D)$ (D random), as in [48], otherwise it is impractical. This is because the integration dimension relies on the dimension of the random space (the total number of input random variables). The Monte Carlo strategy can cope with high uncertainty dimension and large N, albeit at the expense of introducing a statistical error due to sampling, in addition to the bias error. The sampling error is reduced as the number of realizations increases, but at the cost of higher computational burden.

We need to justify that, for each t, $\lim_{N\to\infty} f_{X^N(t)}(x) = f_{X(t)}(x)$, for $x \in \mathbb{R}$. This is a strong mode of convergence. Indeed, as we are working with density functions, almost everywhere convergence on $\mathbb{R}$ implies convergence in $\mathrm{L}^1(\mathbb{R})$: $\|f_{X(t)} - f_{X^N(t)}\|_{\mathrm{L}^1(\mathbb{R})} = \int_{\mathbb{R}} |f_{X(t)}(x) - f_{X^N(t)}(x)|\,\mathrm{d}x \to 0$ as $N \to \infty$. This is due to Scheffé's lemma [170, p. 55], [144]. Convergence in $\mathrm{L}^1(\mathbb{R})$ is also referred to as convergence in total variation [162, p. 41]:

$$\begin{aligned}\left\|\left(\mathbb{P}\circ(X^N(t))^{-1}\right) - \left(\mathbb{P}\circ(X(t))^{-1}\right)\right\|_{\mathrm{TV}} &\coloneqq \sup_{F\in\mathcal{F}} \left|\mathbb{P}[X^N(t)\in F] - \mathbb{P}[X(t)\in F]\right| \\ &= \frac{1}{2}\|f_{X^N(t)} - f_{X(t)}\|_{\mathrm{L}^1(\mathbb{R})}.\end{aligned}$$

It is also equivalent to convergence in terms of the Hellinger distance [84],

$$H\left(\mathbb{P}\circ(X^N(t))^{-1},\mathbb{P}\circ(X(t))^{-1}\right) := \sqrt{\frac{1}{2}\int_{\mathbb{R}}\left(\sqrt{f_{X^N(t)}(x)}-\sqrt{f_{X(t)}(x)}\right)^2\,\mathrm{d}x}$$
$$= \frac{1}{\sqrt{2}}\left\|\sqrt{f_{X^N(t)}}-\sqrt{f_{X(t)}}\right\|_{\mathrm{L}^2(\mathbb{R})},$$

via the elementary inequalities $H^2 \leq \|\cdot\|_{\mathrm{TV}} \leq \sqrt{2}\,H$.

In fact, convergence in $\mathrm{L}^1(\mathbb{R})$ may be generalized to convergence in $\mathrm{L}^p(\mathbb{R})$, for $1<p<\infty$, by imposing boundedness on $\mathbb{R}$ of f_{Y_0}. Indeed, in this case, taking a constant $C>0$ such that $|f_{X^N(t)}(x)|\leq C$ and $|f_{X(t)}(x)|\leq C$, for $N\geq 0$, t and $x\in\mathbb{R}$, the mean value theorem leads to

$$\left|\|f_{X^N(t)}\|^p_{\mathrm{L}^p(\mathbb{R})}-\|f_{X(t)}\|^p_{\mathrm{L}^p(\mathbb{R})}\right| \leq \int_{\mathbb{R}}\left|\left(f_{X^N(t)}(x)\right)^p-\left(f_{X(t)}(x)\right)^p\right|\,\mathrm{d}x$$
$$\leq p\,C^{p-1}\|f_{X^N(t)}-f_{X(t)}\|_{\mathrm{L}^1(\mathbb{R})},$$

therefore $\|f_{X^N(t)}\|_{\mathrm{L}^p(\mathbb{R})}\to\|f_{X(t)}\|_{\mathrm{L}^p(\mathbb{R})}$ as $N\to\infty$. By Scheffé's lemma, there is convergence in $\mathrm{L}^p(\mathbb{R})$: $\|f_{X(t)}-f_{X^N(t)}\|_{\mathrm{L}^p(\mathbb{R})}=(\int_{\mathbb{R}}|f_{X(t)}(x)-f_{X^N(t)}(x)|^p\,\mathrm{d}x)^{1/p}\to 0$ as $N\to\infty$.

The pointwise convergence is the object of the following important Theorem 6.6. The result is proved in the spirit of our contribution [20], by utilizing the concept of Nemytskii operator [20, Remark 2.6], [2, pp. 15–17], [163, pp. 154–163].

Lemma 6.4 *Let $\{V_N\}_{N=1}^{\infty}$ be a sequence of random variables that converges to V in $\mathrm{L}^2(\Omega)$. If $\mathbb{P}[V\in\mathcal{D}_{f_{Y_0}}]=0$, where $\mathcal{D}_{f_{Y_0}}$ is the set of discontinuity points of f_{Y_0}, and if $f_{Y_0}(y)\leq\alpha+\beta y^2$, for certain constants $\alpha,\beta\geq 0$, then $f_{Y_0}(V_N)\to f_{Y_0}(V)$ as $N\to\infty$ in $\mathrm{L}^1(\Omega)$.*

Proof. It suffices to prove that, for every subsequence $\{V_{N_k}\}_{k=1}^{\infty}$, there exists a subsequence $\{V_{N_{k_l}}\}_{l=1}^{\infty}$ such that $\lim_{l\to\infty}f_{Y_0}(V_{N_{k_l}})=f_{Y_0}(V)$ in $\mathrm{L}^1(\Omega)$. Fix any subsequence $\{V_{N_k}\}_{k=1}^{\infty}$. Since $\lim_{k\to\infty}V_{N_k}=V$ in $\mathrm{L}^2(\Omega)$, by [12, Th. 4.9] there exist a subsequence $\{V_{N_{k_l}}\}_{l=1}^{\infty}$ and a random variable $\overline{V}\in\mathrm{L}^2(\Omega)$ such that $\lim_{l\to\infty}V_{N_{k_l}}(\omega)=V(\omega)$ and $|V_{N_{k_l}}(\omega)|\leq\overline{V}(\omega)$ almost surely. Since $\mathbb{P}[V\in\mathcal{D}_{f_{Y_0}}]=0$, the continuous mapping theorem [165, p. 7, Th. 2.3] guarantees that $\lim_{l\to\infty}f_{Y_0}(V_{N_{k_l}}(\omega))=f_{Y_0}(V(\omega))$ almost surely. As $f_{Y_0}(V_{N_{k_l}}(\omega))\leq\alpha+\beta(V_{N_{k_l}}(\omega))^2\leq\alpha+\beta(\overline{V}(\omega))^2\in\mathrm{L}^1(\Omega)$, we can apply the dominated convergence theorem to conclude $\lim_{l\to\infty}f_{Y_0}(V_{N_{k_l}})=f_{Y_0}(V)$ in $\mathrm{L}^1(\Omega)$.

□

Remark 6.5 *As $S_0(t_0) = 1$ and $S_0(t)$ is continuous in $\mathrm{L}^\infty(\Omega)$, we can find a neighborhood of t_0, say $(t_0-\delta, t_0+\delta)$ for certain $\delta > 0$, such that $\|S_0(t)-1\|_\infty < 1/4$ for all $t \in (t_0-\delta, t_0+\delta)$. Hence, $S_0(t) > 3/4 > 0$ almost surely, for $t \in (t_0-\delta, t_0+\delta)$. Notice that δ might not be equal to ∞; for instance, the deterministic function $X(t) = \sin t$ satisfies $\ddot{X}(t)+X(t) = 0$, $X(t_0 = \pi/2) = 1$ and $\dot{X}(t_0 = \pi/2) = 0$.*

For $t \in (t_0-\delta, t_0+\delta)$ fixed, there exists an integer $N_t > 0$ such that $\|S_0^N(t) - S_0(t)\|_\infty < 1/4$, for all $N \geq N_t$. Then $\|S_0^N(t) - 1\|_\infty \leq \|S_0^N(t) - S_0(t)\|_\infty + \|S_0(t) - 1\|_\infty < 1/2$. This implies that $S_0^N(t) > 1/2$ almost surely, $N \geq N_t$.

From now on, we will work with times t in $(t_0-\delta, t_0+\delta)$.

Theorem 6.6 *Suppose the conditions of Theorem 6.3. If f_{Y_0} is continuous on $\mathbb{R}$ and $f_{Y_0}(y) \leq \alpha + \beta y^2$, for certain constants $\alpha, \beta \geq 0$, then we have $\lim_{N\to\infty} f_{X^N(t)}(x) = f_{X(t)}(x)$, for each $t \in (t_0-\delta, t_0+\delta)$ and for every $x \in \mathbb{R}$.*

Proof. Fix $t \in (t_0-\delta, t_0+\delta)$ and $x \in \mathbb{R}$. Let

$$V_N = \frac{x - Y_1 S_1^N(t)}{S_0^N(t)}, \quad V = \frac{x - Y_1 S_1(t)}{S_0(t)} \tag{6.3}$$

(here we drop the explicit dependencies of V_N and V on t and x).
First, notice that $V_N \to V$ as $N \to \infty$ in $\mathrm{L}^2(\Omega)$, as an easy consequence of the following facts: $S_0^N(t) > 1/2$ almost surely, for all $N \geq N_t$, $S_0^N(t) \to S_0(t)$ and $S_1^N(t) \to S_1(t)$ as $N \to \infty$ in $\mathrm{L}^\infty(\Omega)$, and $Y_1 \in \mathrm{L}^2(\Omega)$.
The conditions imposed on f_{Y_0} imply that the Nemytskii operator $V \mapsto f_{Y_0}(V)$ is continuous from $\mathrm{L}^2(\Omega)$ to $\mathrm{L}^1(\Omega)$, by Lemma 6.4.
Hence, $\lim_{N\to\infty} f_{Y_0}(V_N) \to f_{Y_0}(V)$ in $\mathrm{L}^1(\Omega)$. Since $S_0^N(t) > 1/2$ almost surely, for all $N \geq N_t$, and $\lim_{N\to\infty} S_0^N(t) \to S_0(t)$ in $\mathrm{L}^\infty(\Omega)$, we deduce that $f_{Y_0}(V_N)/S_0^N(t) \to f_{Y_0}(V)/S_0(t)$ as $N \to \infty$ in $\mathrm{L}^1(\Omega)$.
In particular, the sequence of expectations, $f_{X^N(t)}(x) = \mathbb{E}[f_{Y_0}(V_N)/S_0^N(t)]$, converges to $f_{X(t)}(x) = \mathbb{E}[f_{Y_0}(V)/S_0(t)]$, which completes the proof.

□

In Section 6.5, the application of Theorem 6.6 will be illustrated numerically on Examples 6.13–6.14.

Remark 6.7 *Having $\lim_{N\to\infty} f_{Y_0}(V_N)/S_0^N(t) = f_{Y_0}(V)/S_0(t)$ in $\mathrm{L}^1(\Omega)$ assures the convergence of the expectations. If convergence of the variances is*

also needed, one needs to extend the convergence to $\mathrm{L}^2(\Omega)$*. In this case, the boundedness condition on* f_{Y_0} *should be* $f_{Y_0}(y) \leq \alpha + \beta|y|$ *(apply an analogous proof to Lemma 6.4).*

Remark 6.8 (Rate of convergence of the density functions) *Note that, under the conditions of Theorem 6.3, if* f_{Y_0} *is Lipschitz continuous on* $\mathbb{R}$ *(this assumption is stronger than the hypotheses of Theorem 6.6), then* $f_{X^N(t)}(x)$ *converges with* N *exponentially to* $f_{X(t)}(x)$*, for* $t \in (t_0 - \delta, t_0 + \delta)$ *and* $x \in \mathbb{R}$*. This is because the Lipschitz condition allows estimating* $|f_{X^N(t)}(x) - f_{X(t)}(x)|$ *via the following inequality:*

$$|f_{X^N(t)}(x) - f_{X(t)}(x)| \leq C_t \left((|x| + 1)\|S_0^N(t) - S_0(t)\|_\infty + \|Y_1\|_2 \|S_1^N(t) - S_1(t)\|_\infty \right),$$

and as discussed in Section 6.2, the Fröbenius method converges exponentially. In the previous expression, C_t *is a constant depending on* t*. Unfortunately, the exponential convergence rate is not uniform with* t *and* x*. As* $|t - t_0|$ *grows, one needs to increase* N *to maintain the accuracy. The same occurs with* $|x|$*, which increases the bias error* $\|S_0^N(t) - S_0(t)\|_\infty$ *linearly.*

In general, if f_{Y_0} *is* γ*-Hölder continuous on* $\mathbb{R}$ *with exponent* $0 < \gamma \leq 1$ *(the case* $\gamma = 1$ *corresponds to Lipschitz continuity), then*

$$\begin{aligned} |f_{X^N(t)}(x) - f_{X(t)}(x)| \leq C_t \big\{ & \|S_0^N(t) - S_0(t)\|_\infty \\ & + \left(|x| \|S_0^N(t) - S_0(t)\|_\infty + \|Y_1\|_2 \|S_1^N(t) - S_1(t)\|_\infty \right)^\gamma \Big\}. \end{aligned}$$

The same conclusion on the convergence holds in this case.

Notice that the regularity of $f_{X^N(t)}(x)$ is inherited from $f_{Y_0}(y)$. These ideas are formalized in the following theorem:

Theorem 6.9 *Under the assumptions of Theorem 6.6, if* f_{Y_0} *is* $C^1(\mathbb{R})$ *with bounded derivative on* $\mathbb{R}$*, then* $f_{X^N(t)}(x)$ *and* $f_{X(t)}(x)$ *are* $C^1(\mathbb{R})$*, with bounded derivatives, and* $\lim_{N\to\infty} f'_{X^N(t)}(x) = f'_{X(t)}(x)$*, for each* $t \in (t_0 - \delta, t_0 + \delta)$ *and for every* $x \in \mathbb{R}$*.*

Proof. Fix $t \in (t_0 - \delta, t_0 + \delta)$. The following facts permit differentiating under the expectation operator that defines $f_{X^N(t)}(x)$ and $f_{X(t)}(x)$ [104, p. 142]: f_{Y_0} is differentiable with bounded derivative, and $S_0^N(t) > 1/2$ almost surely for all $N \geq N_t$. Hence,

$$f'_{X^N(t)}(x) = \mathbb{E}\left[f'_{Y_0}\left(\frac{x - Y_1 S_1^N(t)}{S_0^N(t)} \right) \frac{1}{(S_0^N(t))^2} \right],$$

$$f'_{X(t)}(x) = \mathbb{E}\left[f'_{Y_0}\left(\frac{x - Y_1 S_1(t)}{S_0(t)}\right)\frac{1}{(S_0(t))^2}\right].$$

The continuity and boundedness conditions imposed on f'_{Y_0} entail that the Nemytskii operator $V \mapsto f'_{Y_0}(V)$ is continuous from $\mathrm{L}^2(\Omega)$ to $\mathrm{L}^1(\Omega)$, by Lemma 6.4. Thereby, as in the proof of Theorem 6.6, we deduce that $\lim_{N\to\infty} f'_{X^N(t)}(x) = f'_{X(t)}(x)$, $x \in \mathbb{R}$.

□

Remark 6.10 *It is important to realize that the previous theory works exchanging the role of Y_1 and Y_0. Indeed, even though $S_1(t_0) = 0$, in contrast with $S_0(t_0) = 1$, we do have that $\dot{S}_1(t_0) = 1$. We may choose a neighborhood of t_0, say $(t_0-\mu, t_0+\mu)$ for certain $\mu > 0$, such that $\dot{S}_1(t) > 3/4$ almost surely, for $t \in (t_0-\mu, t_0+\mu)$. We know that, in the sense of $\mathrm{L}^\infty(\Omega)$, $S_1(t) = \int_{t_0}^t \dot{S}_1(r)\,\mathrm{d}r$. Then $|S_1(t)| > \frac{3}{4}|t - t_0| = m_t$ almost surely, for $t \in (t_0-\mu, t_0+\mu)$. In particular, as $m_t > 0$ for $t \in (t_0-\mu, t_0+\mu)\backslash\{t_0\}$, the previous proofs work with Y_1 in lieu of Y_0. The previous theoretical results may be restated in a completely analogous fashion, as*

$$f_{X(t)}(x) = \mathbb{E}\left[f_{Y_1}\left(\frac{x - Y_0 S_0(t)}{S_1(t)}\right)\frac{1}{|S_1(t)|}\right] \tag{6.4}$$

and

$$f_{X^N(t)}(x) = \mathbb{E}\left[f_{Y_1}\left(\frac{x - Y_0 S_0^N(t)}{S_1^N(t)}\right)\frac{1}{|S_1^N(t)|}\right], \tag{6.5}$$

for $t \in (t_0-\mu, t_0+\mu)\backslash\{t_0\}$. In this case, one requires Y_1 to have an absolutely continuous probability law, with density function f_{Y_1}, and to be independent of the rest of the random input parameters in (3.1). For convergence, one imposes continuity for f_{Y_1} on $\mathbb{R}$ and boundedness $f_{Y_1}(y) \le \alpha + \beta y^2$, for certain constants $\alpha, \beta \ge 0$. If f_{Y_1} is Lipschitz continuous on $\mathbb{R}$, then an exponential convergence holds. Finally, if f_{Y_1} is also $C^1(\mathbb{R})$ with bounded derivative on $\mathbb{R}$, then both $f_{X^N(t)}(x)$ and $f_{X(t)}(x)$ are $C^1(\mathbb{R})$, with bounded derivative, and the sequence of derivatives converges. These cases are considered in Example 6.15.

The continuity condition on $\mathbb{R}$ imposed in Theorem 6.6 is somewhat restrictive, as we do not allow some common probability distributions for Y_0 whose density function possesses discontinuity points, such as the Uniform, Exponential or general truncated distributions. Notice that this assumption in Theorem 6.6 may be relaxed to almost everywhere continuity on $\mathbb{R}$, by adding absolute continuity on Y_1. This fact is a consequence of the continuous mapping theorem

[165, p. 7, Th. 2.3]. Indeed, for $t \in (t_0 - \min\{\delta, \mu\}, t_0 + \min\{\delta, \mu\})\backslash\{t_0\}$, as $|S_1(t)| > m_t > 0$ almost surely and Y_1 is absolutely continuous, then $V = \frac{x - Y_1 S_1(t)}{S_0(t)}$ is absolutely continuous, by Lemma 6.2. Therefore, the probability that V lies in the discontinuity set of f_{Y_0} is 0. This assures that $f_{Y_0}(V_N) \to f_{Y_0}(V)$ in $\mathrm{L}^1(\Omega)$ as $N \to \infty$, by Lemma 6.4.

The precise restatement of Theorem 6.6 is the following:

Theorem 6.11 *Suppose the conditions of Theorem 6.3. If f_{Y_0} is almost everywhere continuous on $\mathbb{R}$, $f_{Y_0}(y) \leq \alpha + \beta y^2$ for certain constants $\alpha, \beta \geq 0$, Y_1 is absolutely continuous, and Y_1 is independent of (A, B), then we have $\lim_{N\to\infty} f_{X^N(t)}(x) = f_{X(t)}(x)$, for $t \in (t_0 - \min\{\delta, \mu\}, t_0 + \min\{\delta, \mu\})\backslash\{t_0\}$ and for every $x \in \mathbb{R}$.*

Theorem 6.11 will be applied in Example 6.16. An alternative version, with Y_1 playing the role of Y_0, can be formulated following Remark 6.10. Notice that, nowhere in our theoretical exposition, we require independence between the coefficients of $A(t)$ and $B(t)$. We do not need any assumption on their probability distributions either, which might be discrete or continuous (but always bounded).

The methodology and theory presented in the chapter do not cover all situations. For instance, let us study (3.1) involving discrete uncertainties. Other situations could be analogously analyzed.

Theorem 6.12 *Suppose the conditions of Theorem 6.3. Assume that all the coefficients $A_0, A_1, \ldots, B_0, B_1, \ldots$ are deterministic constants. If f_{Y_0} has at most a countable number of discontinuities on $\mathbb{R}$, $f_{Y_0}(y) \leq \alpha + \beta y^2$ for certain constants $\alpha, \beta \geq 0$, and Y_1 is a discrete random variable, then we have $\lim_{N\to\infty} f_{X^N(t)}(x) = f_{X(t)}(x)$ for almost every $x \in \mathbb{R}$, for each $t \in (t_0 - \delta, t_0 + \delta)$.*

Proof. Fix $t \in (t_0 - \delta, t_0 + \delta)$. Let

$$V_N(x) = \frac{x - Y_1 S_1^N(t)}{S_0^N(t)}, \quad V(x) = \frac{x - Y_1 S_1(t)}{S_0(t)}$$

(now we make the dependence of V_N and V on x explicit). We know that $V_N(x) \to V(x)$ in $\mathrm{L}^2(\Omega)$ as $N \to \infty$, for all $x \in \mathbb{R}$. Given the discontinuity set of f_{Y_0}, $\mathcal{D}_{f_{Y_0}}$, we need to justify that $\mathbb{P}[V(x) \in \mathcal{D}_{f_{Y_0}}] = 0$, for almost every $x \in \mathbb{R}$. In this case, $f_{Y_0}(V_N(x)) \to f_{Y_0}(V(x))$ in $\mathrm{L}^1(\Omega)$ as $N \to \infty$, for almost every $x \in \mathbb{R}$.

Write $\mathcal{D}_{f_{Y_0}} = \{d_1, d_2, d_3, \ldots\}$. As Y_1 is a discrete random variable, its support may be expressed as $\mathcal{S}_{Y_1} = \{y_1^1, y_1^2, y_1^3, \ldots\}$. Then the support of $V(x)$ is $\mathcal{S}_{V(x)} = \{\frac{x - y_1^j S_1(t)}{S_0(t)} : j = 1, 2, 3, \ldots\}$. The problematic x's are those such that $x = y_1^j S_1(t) + d_k S_0(t)$. Let $\Lambda = \{y_1^j S_1(t) + d_k S_0(t) : j, k = 1, 2, 3, \ldots\}$, which is a countable set. For every $x \notin \Lambda$, $\mathbb{P}[V(x) \in \mathcal{D}_{f_{Y_0}}] = 0$. As a consequence, $\lim_{N\to\infty} f_{Y_0}(V_N(x)) = f_{Y_0}(V(x))$ in $\mathrm{L}^1(\Omega)$, $x \notin \Lambda$, by Lemma 6.4. This gives $\lim_{N\to\infty} f_{X^N(t)}(x) = f_{X(t)}(x)$, $x \notin \Lambda$, and we are done.

□

Once again, one can state a similar version with Y_1 playing the role of Y_0 (see Remark 6.10) and working on $(t_0 - \mu, t_0 + \mu)\backslash\{t_0\}$, instead. Example 6.17 covers this situation.

6.4 Computational aspects

We recast the proposed methodology in the form of Algorithm 1, which corresponds to the case of Y_0 having a density, see (6.1); following Remark 6.10, one can exchange the role of Y_0 and Y_1 in Algorithm 1, provided that Y_1 has a density.

By judiciously exploiting its expression in (6.2), $f_{X^N(t)}(x)$ can be approximated via a Monte Carlo procedure [172, pp. 53–54] to evaluate the expectation: using M randomly generated realizations of Y_1, $S_0^N(t)$ and $S_1^N(t)$, we compute the sample average of $V_N(x,t)$ in (6.3). Algorithm 1 corresponds to symbolic computations with symbolic variables x and t [73]; it computes a function $f_X^{N,M}(x,t)$, which is a complex closed-form expression approximating $f_{X(t)}(x)$. To speed up the execution of the algorithm, numerical values of t and/or x may be provided.

The estimation error can be split into two contributions: $f_{X(t)}(x) - f_X^{N,M}(x,t) = \theta_N(x,t) + \mathcal{E}_{N,M}(x,t)$. The first contribution, $\theta_N(x,t) = f_{X(t)}(x) - f_{X^N(t)}(x)$, is the bias error caused by the truncation order N in the Fröbenius method. It is deterministic and decays exponentially as $N \to \infty$ for each t and x by Remark 6.8. The second contribution is the sampling error $\mathcal{E}_{N,M}(x,t) = f_{X^N(t)}(x) - f_X^{N,M}(x,t)$, due to using a finite number M of samples (statistical error). This contribution is random and $\mathcal{E}_{N,M}(x,t) \to 0$ with M almost surely, as a consequence of the law of large numbers. If the variance

$$\sigma_N^2(x,t) \coloneqq \mathbb{V}\left[f_{Y_0}\left(\frac{x - Y_1 S_1^N(t)}{S_0^N(t)}\right)\frac{1}{S_0^N(t)}\right] \tag{6.6}$$

Algorithm 1 Estimation of $f_{X^{N}(t)}(x)$ via a classical Monte Carlo procedure.

Inputs: t_0; N; f_{Y_0}; probability distribution of $A_0, \ldots, A_N$, $B_0, \ldots, B_N$, Y_1; and number M of realizations in the classical Monte Carlo procedure. Here, t and x will be symbolic variables.

Required: Hypotheses of Theorem 6.6 or Theorem 6.11.

$S_{0,0} \leftarrow 1$, $S_{0,1} \leftarrow 0$, $S_{1,0} \leftarrow 0$, $S_{1,1} \leftarrow 1$ ▷ Initial conditions

$\Sigma \leftarrow 0$ ▷ Initialize the samples sum

for $i = 1, \ldots, M$ **do** ▷ Monte Carlo loop

  Draw randomly a realization of $(A_0, \ldots, A_{N-2}, B_0, \ldots, B_{N-2})$ and Y_1

  for $n = 0, \ldots, N-2$ **do**

    $S_{0,n+2} \leftarrow \frac{-1}{(n+2)(n+1)} \sum_{m=0}^{n} [(m+1)A_{n-m}S_{0,m+1} + B_{n-m}S_{0,m}]$

    $S_{1,n+2} \leftarrow \frac{-1}{(n+2)(n+1)} \sum_{m=0}^{n} [(m+1)A_{n-m}S_{1,m+1} + B_{n-m}S_{1,m}]$

  end for

  $S_0^N(t) \leftarrow 1 + \sum_{n=1}^{N} S_{0,n}(t-t_0)^n$ ▷ Realization of $S_0^N(t)$

  $S_1^N(t) \leftarrow \sum_{n=1}^{N} S_{1,n}(t-t_0)^n$ ▷ Realization of $S_1^N(t)$

  $\Sigma \leftarrow \Sigma + f_{Y_0}\left(\frac{x - Y_1 S_1^N(t)}{S_0^N(t)}\right) \frac{1}{|S_0^N(t)|}$ ▷ Update the samples sum

end for

$f_X^{N,M}(x,t) \leftarrow \Sigma / M$ ▷ Set sample average

Return $f_X^{N,M}(x,t)$ ▷ Approximation of $f_{X^{N}(t)}(x)$

is finite, then the asymptotic probability distribution of $\mathcal{E}_{N,M}(x,t)$ as $M \to \infty$ is, by the central limit theorem, Normal$(0, \sigma_N^2(x,t)/M)$. The variance $\sigma_N^2(x,t)$ tends, as $N \to \infty$, to $\sigma^2(x,t) := \mathbb{V}[f_{Y_0}(\frac{x-Y_1 S_1(t)}{S_0(t)})\frac{1}{S_0(t)}]$ (see Remark 6.7). In this case, we say that the sampling error is of order $1/\sqrt{M}$, and write $\mathcal{O}(1/\sqrt{M})$. On the contrary, if $\sigma_N^2(x,t) = \infty$, then the almost sure convergence $\mathcal{E}_{N,M}(x,t) \to 0$ with M remains valid, although it might be much slower and affect the approximation to $f_{X^N(t)}(x)$ severely. See the forthcoming Example 6.15 for an illustration of this issue.

Even though the bias error decays very fast, the sampling error is inevitable. In numerical computations, for fixed M, there is usually an index N from which the global error does not go down anymore because the sampling error $\mathcal{O}(1/\sqrt{M})$ becomes dominant.

Within the main loop of Algorithm 1 (loop over the samples), we first generate one realization for each random variable $A_0, \ldots, A_{N-2}$, $B_0, \ldots, B_{N-2}$ and Y_1; these realizations are used to compute by recursion the corresponding realizations of $S_0^N(t)$ and $S_1^N(t)$. In our implementation, this procedure is more effective than expressing first $S_0^N(t)$ and $S_1^N(t)$ recursively in terms of symbolic variables $A_0, \ldots, A_{N-2}$, $B_0, \ldots, B_{N-2}$ and Y_1, and then evaluate for the realizations of $A_0, \ldots, A_{N-2}$, $B_0, \ldots, B_{N-2}$ and Y_1. This is due to the excessive complexity of the symbolic expressions of $S_0^N(t)$ and $S_1^N(t)$, which makes the computational time of their evaluation for specific realizations prohibitively large.

The computational complexity of Algorithm 1 is at most $\mathcal{O}(MN^2)$ (the nested loop over n demands $\sum_{n=0}^{N-2} \mathcal{O}(n) = \mathcal{O}(N^2)$ operations in general). As we show in the following Section 6.5, the implemented algorithm is certainly applicable and suitable for stochastic computations.

By taking $M = \mathcal{O}(1/\epsilon^2)$, the variance of the statistical error is $\mathbb{V}[\mathcal{E}_{N,M}(x,t)] = \mathcal{O}(\epsilon^2)$ (assuming the variance in (6.6) finite). Under exponential convergence of the bias, by picking $N = \mathcal{O}(\log(1/\epsilon)) + \mathcal{O}(1)$ the bias error is $|\theta_N(x,t)| = \mathcal{O}(\epsilon)$. Then the root mean square error of the algorithm is $\|f_{X(t)}(x) - f_X^{N,M}(x,t)\|_2 = \sqrt{\theta_N(x,t)^2 + \mathbb{V}[\mathcal{E}_{N,M}(x,t)]} = \mathcal{O}(\epsilon)$, with a computational complexity $\mathcal{O}(MN^2) = \mathcal{O}\left(\epsilon^{-2}\log^2\epsilon\right)$.

The complexity of Algorithm 1 is significantly reduced if $A(t)$ and $B(t)$ are random polynomials, instead of infinite series. Suppose for instance that $A_j = 0$ and $B_j = 0$, for $j \geq N_0 - 1$. Then, within the nested loop over n, we actually sum N_0 terms, instead of n terms. Therefore, the nested loop demands

$N_0\mathcal{O}(N) = \mathcal{O}(N)$ operations. The whole algorithm then requires $\mathcal{O}(MN)$ operations only. If we take $M = \mathcal{O}(1/\epsilon^2)$ and $N = \mathcal{O}(\log(1/\epsilon)) + \mathcal{O}(1)$ to ensure a root mean square error of order ϵ, the computational complexity becomes $\mathcal{O}(MN) = \mathcal{O}(\epsilon^{-2}\log(\epsilon^{-1}))$.

In the case in which $A(t)$ and $B(t)$ are deterministic expansions, the loop over n and the assignments for $S_0^N(t)$ and $S_1^N(t)$ may be run once for all at the beginning of the algorithm and before the loop over the samples. The computational complexity then reduces even more to $\mathcal{O}(M)+\mathcal{O}(N^2)$ operations and the global cost is generally dominated by the sampling. To guarantee a global root mean square error of order ϵ with $M = \mathcal{O}(1/\epsilon^2)$ and $N = \mathcal{O}(\log(1/\epsilon)) + \mathcal{O}(1)$, the computational complexity becomes $\mathcal{O}(\epsilon^{-2})+\mathcal{O}(\log^2\epsilon) = \mathcal{O}(\epsilon^{-2})$. This scenario allows increasing M and obtaining more accurate results by improving the statistical convergence. If $A(t)$ and $B(t)$ are simply deterministic polynomials, then the overall cost reduces further to $\mathcal{O}(M) + \mathcal{O}(N)$ operations, which yields in the end $\mathcal{O}(\epsilon^{-2})$ calculations.

In the view of computational applications, an important drawback of our exposition is the lack of awareness on the specific values of δ and μ, which are necessary to prove the theoretical convergence. Given any t, one can apply Algorithm 1 and check the convergence of the estimator with M and N. The results can be validated using other stochastic methods and using statistics based on the estimated density. Notice that, in Algorithm 1, we have put $|S_0^N(t)|$ instead of $S_0^N(t)$. Even though we assume that $S_0^N(t) > 0$ almost surely, for $t \in (t_0 - \delta, t_0 + \delta)$ and $N \geq N_t$, the absolute value ensures positiveness in numerical applications even if $|t - t_0| \geq \delta$.

6.5 Numerical examples

In this section, we numerically illustrate our theoretical findings, using Algorithm 1 to estimate the density of the solution to (3.1). Several cases, differing by the probability distributions of the random input coefficients, are considered to cover a large class of situations and show the broad applicability of our theory.

In each of these examples, we first check that the necessary theoretical conditions hold; we then estimate the density function $f_{X^N(t)}(x)$ for several increasing values of N to highlight the convergence toward $f_{X(t)}(x)$. To this end, we employ the symbolic Monte Carlo sampling procedure outlined in Algorithm 1.

The theoretical results of Section 6.3 motivate the structure and the choice of the following five examples. In Example 6.13, we address the case where $A(t)$ and $B(t)$ are random polynomials; while Example 6.14 concerns infinite expansions and infinite dimensionality. These first two examples showcase the applicability of Theorem 6.6. Example 6.15 is designed to highlight Remark 6.10. Up to this example, f_{Y_0} or f_{Y_1} are continuous on the whole real line. In contrast, Example 6.16 considers experiments with f_{Y_0} possessing discontinuity points, thus evoking Theorem 6.11. Finally, Example 6.17 considers the case where $A(t)$ and $B(t)$ are deterministic, so that Theorem 6.12 applies.

The implementations and computations are performed with Mathematica®, version 11.2, owing to its capability to handle symbolic computations. In general, Algorithm 1 is applied with $M = 20,000$ samples, as beyond this limit, the computational burden is becoming massive. The output function $f_X^{N,M=20,000}(x,t)$ is handled symbolically on t and x. To simplify the notations, we refer to the Monte Carlo estimate $f_X^{N,M=20,000}(x,t)$ as $\hat{f}_{X^N(t)}(x)$. We recall that the estimate $\hat{f}_{X^N(t)}(x)$ has two sources of error: bias and sampling. Although the bias error decays very fast (exponentially under the conditions of Remark 6.8), the sampling error is unavoidable and at least of order $\mathcal{O}(1/\sqrt{M})$.

In each one of the following examples, we perform a complete analysis of the errors. As the exact density function $f_{X(t)}(x)$ is not known, we first analyze differences in consecutive (in N) estimates, both pointwise, using

$$\delta\epsilon^N(x,t) \coloneqq |\hat{f}_{X^{N+1}(t)}(x) - \hat{f}_{X^N(t)}(x)|, \tag{6.7}$$

and globally, using the norm

$$\Delta\epsilon^N(t) \coloneqq \|\hat{f}_{X^{N+1}(t)} - \hat{f}_{X^N(t)}\|_{\mathrm{L}^1(\mathbb{R})}. \tag{6.8}$$

As successive differences do not directly characterize the error, we also report

$$E^N(t) \coloneqq \|\hat{f}_{X^L(t)} - \hat{f}_{X^N(t)}\|_{\mathrm{L}^1(\mathbb{R})} \tag{6.9}$$

for some pre-fixed $L \gg 1$, selected such that $\hat{f}_{X^L(t)}$ plays the role of a bias-free estimate of function $f_{X(t)}$. We set $L = 30$ in the following. The $\mathrm{L}^1(\mathbb{R})$ norms are computed by direct numerical integration, using a standard quadrature rule (standard `NIntegrate` routine in Mathematica®).

	$N=1$	$N=2$	$N=3$	$N=4$	$N=5$
$t=0.5$	0.903091	0.622968	0.270690	0.0923362	0.0178834
	$N=6$	$N=7$	$N=8$	$N=9$	$N=10$
$t=1$	0.691809	0.263246	0.0912177	0.0345686	0.026688
	$N=11$	$N=12$	$N=13$	$N=14$	$N=15$
$t=1.5$	0.348643	0.180075	0.0721679	0.0320314	0.0198364

Table 6.1: Norm $\Delta\epsilon^N(t)$ of differences in consecutive estimates (see (6.8)) for different times t and truncation orders N. This table corresponds to Example 6.13.

Example 6.13 We start with the following stochastic problem where both $A(t)$ and $B(t)$ are random polynomials of degree 1:

$$\begin{cases} \ddot{X}(t) + (A_0 + A_1 t)\dot{X}(t) + (B_0 + B_1 t)X(t) = 0,\ t \in \mathbb{R}, \\ X(t_0 = 0) = Y_0, \\ \dot{X}(t_0 = 0) = Y_1, \end{cases} \tag{6.10}$$

where $A_0 = 4$, $A_1 \sim \text{Uniform}(0,1)$, $B_0 \sim \text{Gamma}(2,2)$, $B_1 \sim \text{Bernoulli}(0.35)$, $Y_0 \sim \text{Normal}(2,1)$ and $Y_1 \sim \text{Poisson}(2)$ are independent random variables. In order for the hypotheses of Theorem 3.4 and Theorem 6.6 to be satisfied, the Gamma distribution is truncated. For the Gamma distribution with shape and rate 2, it can be checked that the interval $[0,4]$ contains approximately 99.7% of the probability, so we actually consider $B_0 \sim \text{Gamma}(2,2)|_{[0,4]}$.

By Theorem 3.4, the unique mean square solution to (6.10) can be written as a random power series $X(t) = \sum_{n=0}^{\infty} X_n t^n$ that is mean square convergent for all $t \in \mathbb{R}$. With Theorem 6.6, we approximate pointwise the probability density function $f_{X(t)}(x)$ with $\hat{f}_{X^N(t)}(x)$, $N \geq 0$, and use Algorithm 1 taking advantage from the fact that $A(t)$ and $B(t)$ are random polynomials and not infinite expansions.

We consider times $t = 0.5$, 1 and 1.5. In Figure 6.1 we present the graphs of $\hat{f}_{X^N(t)}(x)$ at the corresponding times. Observe that the estimates are smooth, due to the regularity of the initial density f_{Y_0}, see Theorem 6.9. Observe also that, as N increases, the density functions become closer, reflecting the theoretical convergence. The convergence is made clear in the corresponding successive differences $\delta\epsilon^N(x,t)$ (see (6.7)) reported in Figure 6.2. Table 6.1 presents the $\text{L}^1(\mathbb{R})$ norms of the successive differences, $\Delta\epsilon^N(t)$ (see (6.8)).

Figure 6.3 reports (in log-scale) the error estimate $E^N(t)$ (see (6.9)), for $t = 0.5$, $t = 1$ and $t = 1.5$. From the plot, it is clear that there is an index N from

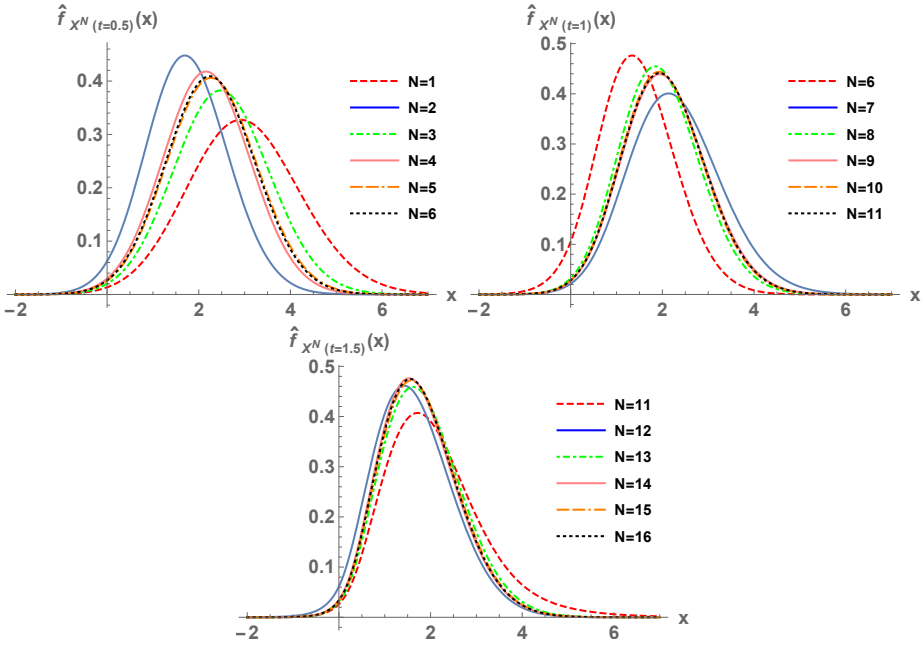

Figure 6.1: Graphical representations of the Monte Carlo estimates $\hat{f}_{X^N(t)}(x)$ at $t = 0.5$, $t = 1$ and $t = 1.5$, with orders of truncation $N = 1$–6, $N = 6$–11 and $N = 11$–16, respectively. This figure corresponds to Example 6.13.

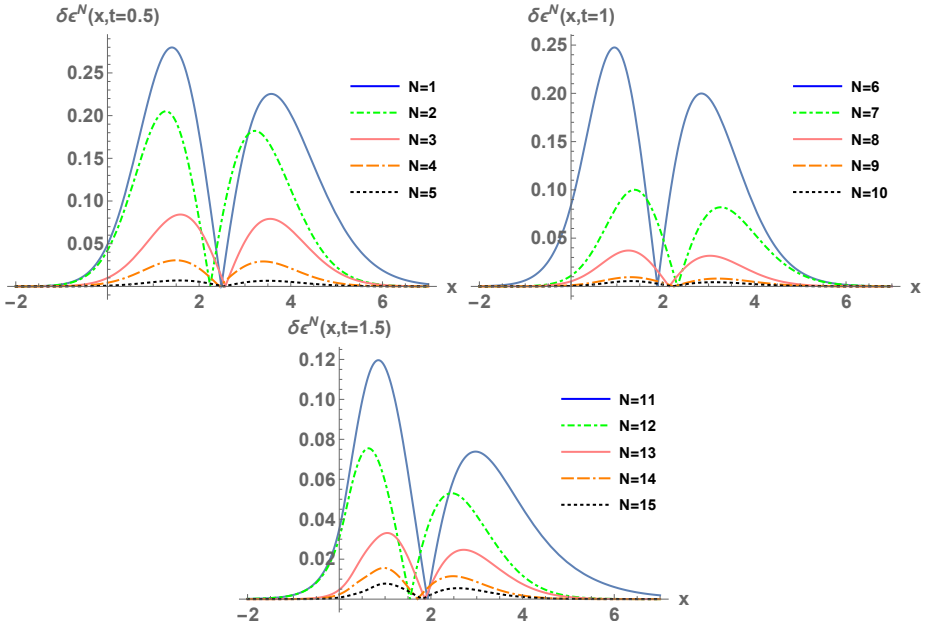

Figure 6.2: Differences in consecutive estimates $\delta\epsilon^N(x,t)$ (see (6.7)) at $t = 0.5$, $t = 1$ and $t = 1.5$, with orders of truncation $N = 1$–5, $N = 6$–10 and $N = 11$–15, respectively. This figure corresponds to Example 6.13.

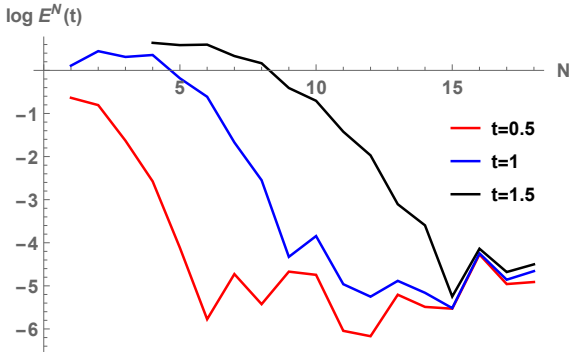

Figure 6.3: Error $E^N(t)$ (see (6.9)), for different times as indicated. This figure corresponds to Example 6.13.

which the error does not go down anymore, because of the sampling error (recall that we used a fixed number of samples $M = 20,000$). Notice also that, as $|t - t_0| = |t|$ gets larger, a higher order of truncation N is required to enhance the approximations of $f_{X(t)}(x)$.

In Figure 6.4, we report the error estimate, $E^N(t)$, as a function of the consecutive difference, $\Delta\epsilon^N(t)$, for $t = 0.5$, $t = 1$ and $t = 1.5$. We also plot a regression line through the data at a given t,

$$\log E^N(t) \approx \log \beta(t) + \alpha(t) \log \Delta\epsilon^N(t), \tag{6.11}$$

to reflect the exponential relationship

$$E^N(t) \approx \beta(t) \left(\Delta\epsilon^N(t)\right)^{\alpha(t)}. \tag{6.12}$$

There are three regression lines, one for each time t. We observe a strong linear relation with N between the errors and the successive differences in log-scale, with slope $\alpha(t)$ being approximately 1, at least up to the truncation order at which the sampling error becomes dominant. This finding suggests that it is possible to estimate the norm of the bias error, $\|\theta_N(\cdot, t)\|_{\mathrm{L}^1(\mathbb{R})}$, from the norm of the successive differences $\Delta\epsilon^N(t)$, provided that M is large enough, and choose N according to the targeted accuracy.

Example 6.14 In the second example, we consider

$$\begin{cases} \ddot{X}(t) + A(t)\dot{X}(t) + B(t)X(t) = 0, \ t \in (-1, 1), \\ X(t_0 = 0) = Y_0, \\ \dot{X}(t_0 = 0) = Y_1, \end{cases} \tag{6.13}$$

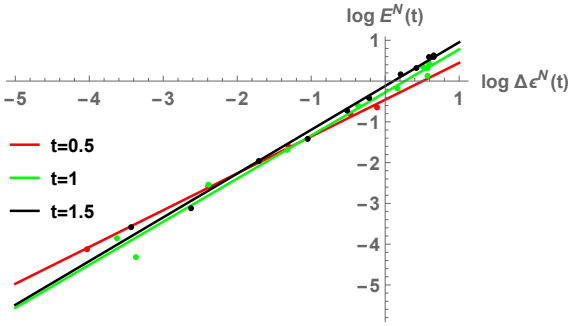

Figure 6.4: Relation between $\log E^N(t)$ and $\log \Delta\epsilon^N(t)$, for $t = 0.5$, $t = 1$ and $t = 1.5$. Also reported are linear regressions. This figure corresponds to Example 6.13.

$A(t)$ and $B(t)$ having infinite expansions with coefficients $A_n \sim \text{Beta}(11, 15)$ for $n \geq 0$, $B_0 = 0$, $B_n = 1/n^2$ for $n \geq 1$, while $Y_0 \sim f_{Y_0}(y) = \frac{\sqrt{2}}{\pi(1+y^4)}$ $(-\infty < y < \infty)$ and $Y_1 \sim \text{Poisson}(2)$. All these random quantities are assumed to be independent. The power series of $A(t)$ and $B(t)$ converge on $(-1, 1)$ (that is for $r = 1$), so the mean square solution $X(t) = \sum_{n=0}^{\infty} X_n t^n$ given by Theorem 3.4 is defined on $(-1, 1)$. Theorem 6.6 allows approximating $f_{X(t)}(x)$ with $\hat{f}_{X^N(t)}(x)$, $N \geq 0$.

Figure 6.5 shows graphical representations of $\hat{f}_{X^N(t)}(x)$ for times $t = 0.25$, 0.5, 0.75 and 0.99, with orders of truncation $N = 1$–5. The evident regularity of $\hat{f}_{X^N(t)}(x)$ is inherited from the smoothness of the density f_{Y_0}, by Theorem 6.9.

To better assess the convergence, Figure 6.6 shows the successive differences $\delta\epsilon^N(x, t)$ defined in (6.7) at the same times as in Figure 6.5; these differences are decreasing to 0 pointwise as theoretically expected, see Theorem 6.6. As pointwise convergence of densities implies $\text{L}^1(\mathbb{R})$ convergence, we report in Table 6.2 the consecutive norms $\Delta\epsilon^N(t)$ defined by (6.8). The norms decay, albeit not monotonically; for instance, when $t = 0.25$ the difference is larger for $N = 4$ than in $N = 3$.

Figure 6.7 reports the error estimates $\log E^N(t)$ defined in (6.9). We see that the errors decrease quickly before stagnating because of the sampling error. This example, despite being more complex than the previous one in Example 6.13, in terms of dimensionality, requires smaller orders N, since for $t \in (-1, 1)$ we have $|t - t_0| = |t| < 1$, which implies $|t - t_0|^n \overset{n\to\infty}{\longrightarrow} 0$.

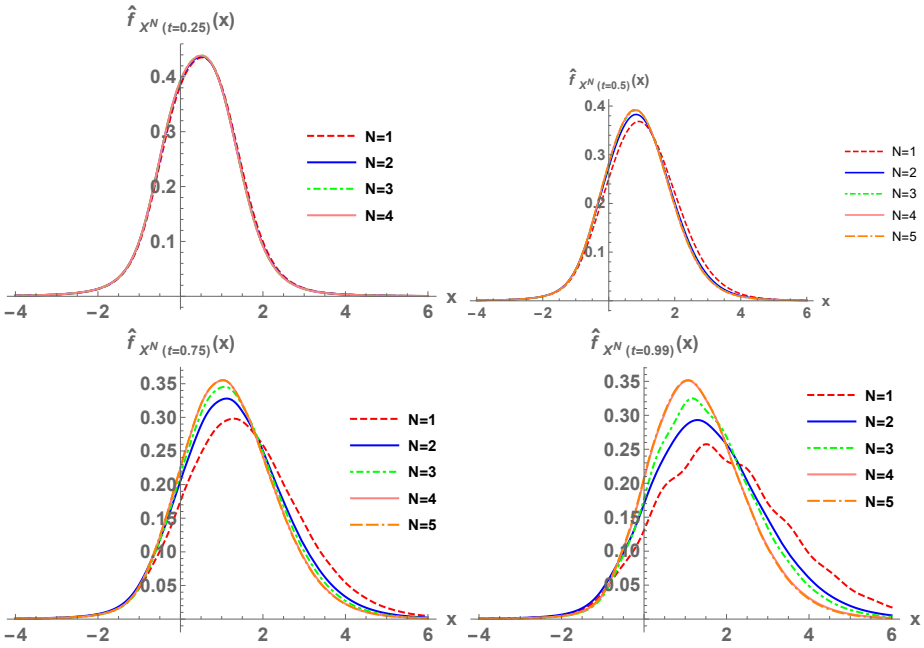

Figure 6.5: Graphical representations of the Monte Carlo estimates $\hat{f}_{X^N(t)}(x)$ at $t = 0.25$, $t = 0.5$, $t = 0.75$ and $t = 0.99$, with orders of truncation N as indicated. This figure corresponds to Example 6.14.

	$N = 1$	$N = 2$	$N = 3$	$N = 4$
$t = 0.25$	0.0215530	0.00607417	0.000600201	0.00167170
$t = 0.5$	0.0776298	0.0226985	0.00913760	0.00318041
$t = 0.75$	0.147952	0.0545970	0.0436801	0.00419704
$t = 0.99$	0.225868	0.0945261	0.127495	0.00985133

Table 6.2: Norm $\Delta\epsilon^N(t)$ of differences in consecutive estimates (see (6.8)) for different times t and truncation orders N. This table corresponds to Example 6.14.

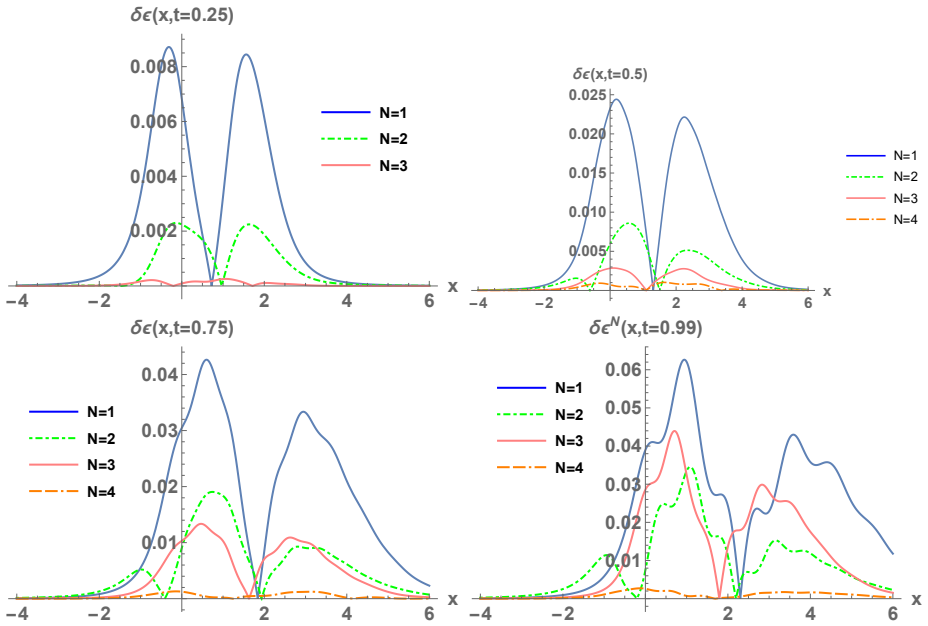

Figure 6.6: Differences in consecutive estimates $\delta\epsilon^N(x,t)$ (see (6.7)) at $t = 0.25$, $t = 0.5$, $t = 0.75$ and $t = 0.99$, and for orders of truncation as indicated. The plots correspond to Example 6.14.

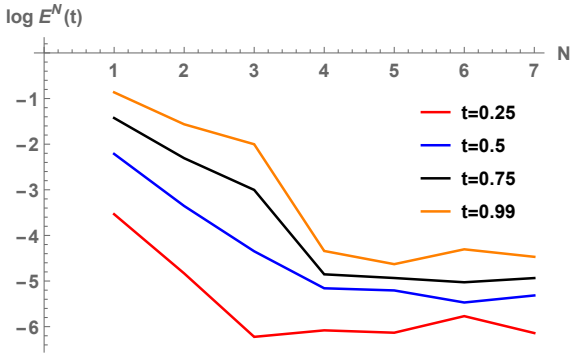

Figure 6.7: Error $E^N(t)$ in (6.9), for different times as indicated. This figure corresponds to Example 6.14.

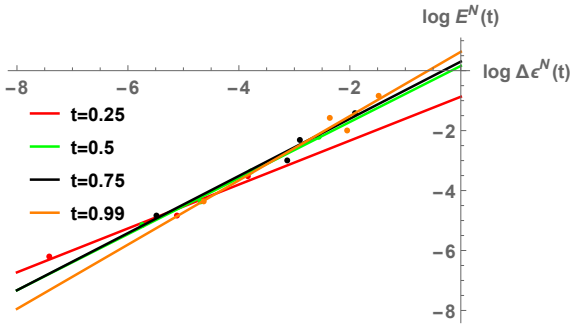

Figure 6.8: Relation between $\log E^N(t)$ and $\log \Delta\epsilon^N(t)$, for $t = 0.25$, 0.5, 0.75 and 0.99. Also reported are linear regressions. This figure corresponds to Example 6.14.

Figure 6.8 aims at showing the relation between the errors $E^N(t)$ and the successive differences $\Delta\epsilon^N(t)$. Specifically, for the times t shown, a collinearity is found in log-scale through the models (6.11)–(6.12). In other words, the decay pattern of the consecutive differences characterizes the convergence of the global error as long as the bias error dominates the sampling error.

Example 6.15 In this example, we consider degree one polynomial problem in (6.10), with the following independent probability distributions: $A_0 = 4$, $A_1 \sim \text{Uniform}(0,1)$, $B_0 \sim \text{Gamma}(2,2)|_{[0,4]}$, $B_1 \sim \text{Bernoulli}(0.35)$, $Y_0 \sim \text{Poisson}(2)$ and $Y_1 \sim \text{Normal}(2,1)$. This example coincides with Example 6.13, except that Y_0 and Y_1 have been interchanged: now Y_0 is discrete, while Y_1 is absolutely continuous. This exchange puts this example in a different theoretical case compared to Example 6.13.

By Theorem 3.4, the unique mean square solution is expressible as a random power series $X(t) = \sum_{n=0}^{\infty} X_n t^n$ that is mean square convergent for all $t \in \mathbb{R}$. According to Remark 6.10, we can approximate the probability density function of $X(t)$, $f_{X(t)}(x)$, for $t \neq 0$.

We work at times $t = 0.5$, 1 and 1.5. Figure 6.9 reports the approximations $\hat{f}_{X^N(t)}(x)$. As N grows, the graphical representations tend to overlap, denoting the convergence of the expansions. The densities are all smooth, as expected from the smoothness of f_{Y_1}, except for $\hat{f}_{X^{N=12}(t=1.5)}(x)$ whose estimate presents noisy features.

The noisy features in $\hat{f}_{X^{N=12}(t=1.5)}(x)$ are due to several reasons. First, there is a computational issue of Mathematica® caused by numerical overflow-underflow

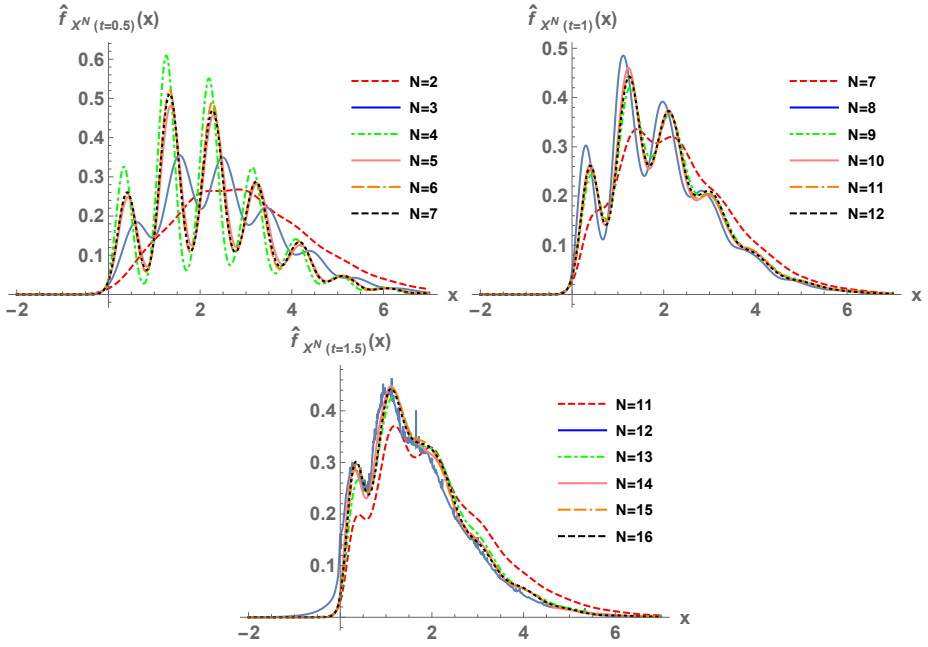

Figure 6.9: Graphical representations of the Monte Carlo estimates $\hat{f}_{X^N(t)}(x)$ at $t = 0.5$, $t = 1$ and $t = 1.5$, with varying orders of truncation N as indicated. This figure corresponds to Example 6.15.

	$N = 2$	$N = 3$	$N = 4$	$N = 5$	$N = 6$
$t = 0.5$	0.320171	0.618382	0.333094	0.0893759	0.0291202
	$N = 7$	$N = 8$	$N = 9$	$N = 10$	$N = 11$
$t = 1$	0.308293	0.185148	0.0605758	0.0256469	0.0216692
	$N = 11$	$N = 12$	$N = 13$	$N = 14$	$N = 15$
$t = 1.5$	0.301694	0.155185	0.0677696	0.0375570	0.0202081

Table 6.3: Norm $\Delta\epsilon^N(t)$ of differences in consecutive estimates (see (6.8)) for different times t and truncation orders N. This table corresponds to Example 6.15.

when too small or too large quantities are involved (for instance $\exp(z)$ for $|z| \gg 1$). Some sample paths of $S_1^{N=12}(t)$ are vanishing near $t = 1.5$, thus making the denominator $S_1^{N=12}(t)$ in the definition of $V_N(t)$ (in (6.3) by for the role of Y_0 and Y_1 exchanged) very small, with a loss of precision as a result. This is illustrated in Figure 6.10, where we show some randomly generated sample paths of $S_1^{N=12}(t)$. We also report sample paths of $S_1^N(t)$ for $N \in \{11, 13, 14\}$ for comparison. Second, and not totally unrelated to the numerical overflow, we have $\mathbb{V}[1/|S_1^N(t = 1.5)|] = \infty$ for $N = 12$ when it remains finite for $N \in \{11, 13, 14\}$. As a result, the variance $\sigma^2_{N=12}$ in the Monte Carlo method (see (6.6)) is unbounded or very large for $N = 12$, while it behaves well for $N \in \{11, 13, 14\}$ as shown in Figure 6.11. As a result, for $N = 12$, the convergence of the Monte Carlo procedure is slowed down due to the large or infinite variance, the rate $\mathcal{O}(1/\sqrt{M})$ is not obtained (see the discussion from Section 6.4), and some noisy features plague the estimator.

Luckily, the noise in $\hat{f}_{X^{N=12}(t=1.5)}(x)$ is not present for $N > 12$. In situations where large or infinite variance occurs for some N, one should focus on the truncation orders N for which the approximation $\hat{f}_{X^N(t)}(x)$ behaves nicely, without noise. In this manner, correct approximations to $f_{X(t)}(x)$ are obtained with a feasible number of samples.

Figure 6.12 presents the consecutive differences $\delta\epsilon^N(x, t)$ given by (6.7). These consecutive differences are not monotonically decreasing with N, although a decay pattern towards 0 is perceptible. Further, the impact of the noisy estimate $\hat{f}_{X^{N=12}(t=1.5)}(x)$ is clearly visible in the reported differences. Besides, the plots are entirely consistent with the theoretical results and Remark 6.10. In Table 6.3, we report the corresponding $\mathrm{L}^1(\mathbb{R})$ norms $\Delta\epsilon^N(t)$ (see (6.8)) as a summary of Figure 6.12.

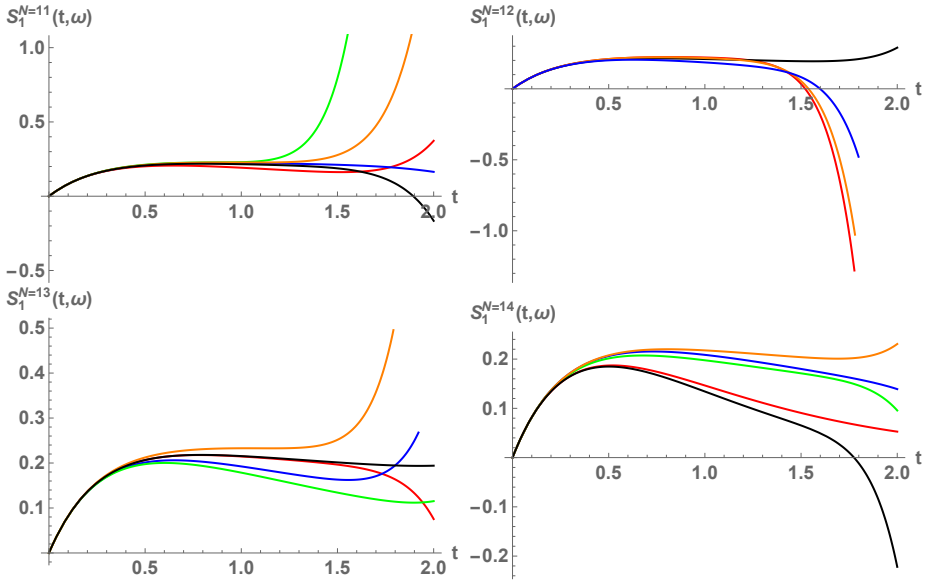

Figure 6.10: Random trajectories of $S_1^N(t)$ for $N = 11$–14. For $N = 12$, observe that some trajectories vanish very close to $t = 1.5$, while for $N \neq 12$ the trajectories remain away from 0. This figure corresponds to Example 6.15.

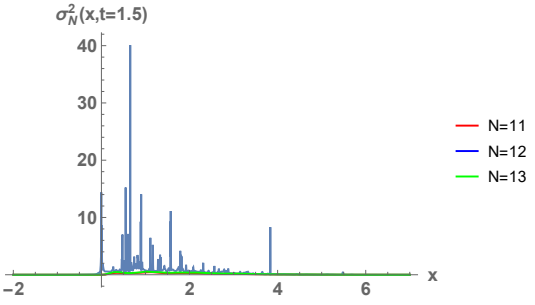

Figure 6.11: Plot of $\sigma_N^2(x, t = 1.5)$, for $N = 11$–13. This figure corresponds to Example 6.15.

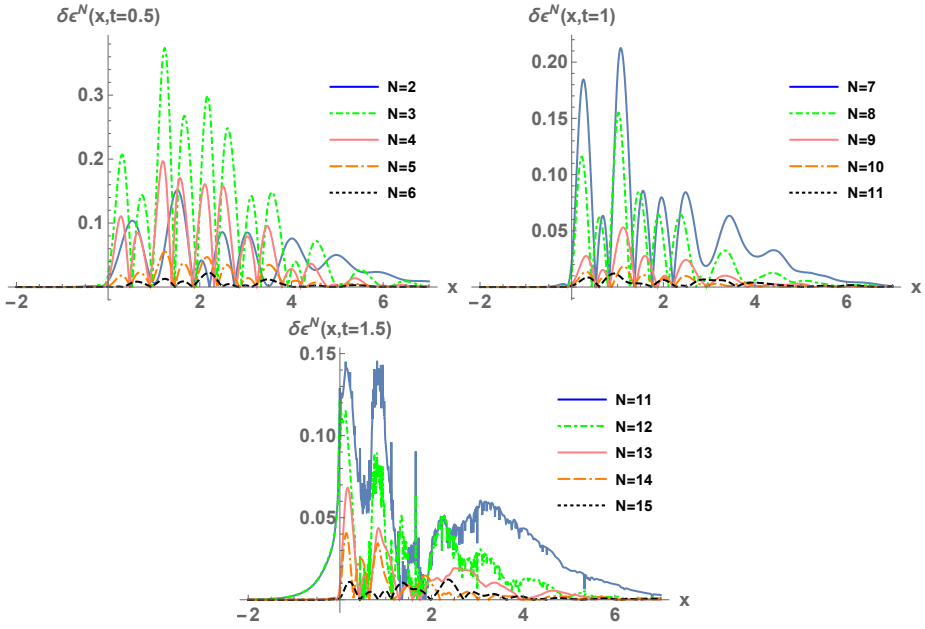

Figure 6.12: Differences in consecutive estimates $\delta\epsilon^N(x,t)$ at $t = 0.5$, $t = 1$ and $t = 1.5$, with orders of truncation as indicated. This figure corresponds to Example 6.15.

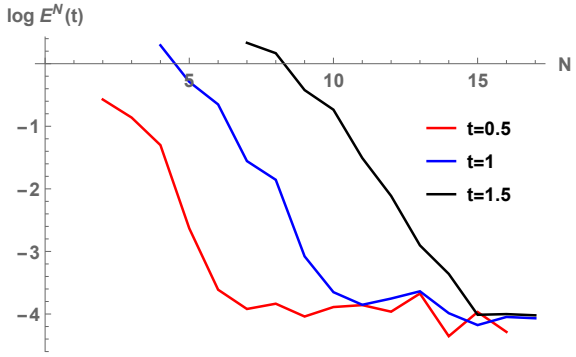

Figure 6.13: Error $E^N(t)$ (see (6.9)), for different times as indicated. This figure corresponds to Example 6.15.

Figure 6.13 reports the estimate errors $\log E^N(t)$ in (6.9). Again, the convergence and the sampling error are observed. This example also emphasizes that the Fröbenius method deteriorates for large times, as N needs to increase with t to maintain accurate approximations.

Example 6.16 Consider the problem (6.13), with $A_n \sim \text{Beta}(11, 15)$ for $n \geq 0$, $B_0 = 0$, $B_n = 1/n^2$ for $n \geq 1$, $Y_0 \sim \text{Uniform}(-1, 1)$ and $Y_1 \sim \text{Exponential}(2)$. These random inputs are again assumed to be independent. In contrast with Example 6.14, the probability density function of Y_0 has now two discontinuity points at $y_0 = \pm 1$, while Y_1 follows an absolutely continuous law. Hence, Theorem 6.6 cannot be employed here.

The mean square analytic solution $X(t) = \sum_{n=0}^{\infty} X_n t^n$ given by Theorem 3.4 is defined on $(-1, 1)$ and we must apply Theorem 6.11 to approximate $f_{X(t)}(x)$ for $t \neq 0$. We compute the approximations at times $t = 0.25$, 0.5, 0.75 and 0.99, with orders of truncation $N = 1$–5 for all the three times. Figure 6.14 depicts the graphs of $\hat{f}_{X^N(t)}(x)$. Promptly, the successive approximations of the density function tend to superimpose, thus entailing rapid convergence to the target density function $f_{X(t)}(x)$. In contrast to Example 6.14, the non-differentiability of the approximated density functions inherited from f_{Y_0} is evident (here one cannot apply Theorem 6.9). Thereby, our method can capture peaks induced by non-differentiability. This feature is a definite advantage of our method, compared to classical sample paths approximation methods where a kernel density estimation of the density would smear-out the approximation at the non-differentiability points.

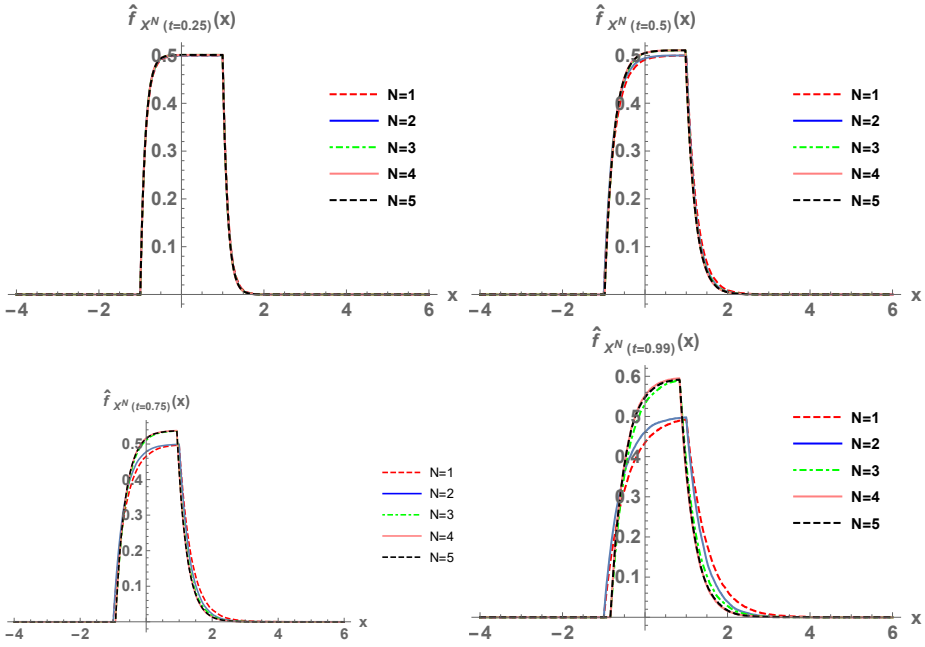

Figure 6.14: Graphical representations of the Monte Carlo estimates $\hat{f}_{X^N(t)}(x)$ at $t = 0.25$, $t = 0.5$, $t = 0.75$ and $t = 0.99$, with orders of truncation $N = 1$–5 in all cases. This figure corresponds to Example 6.16.

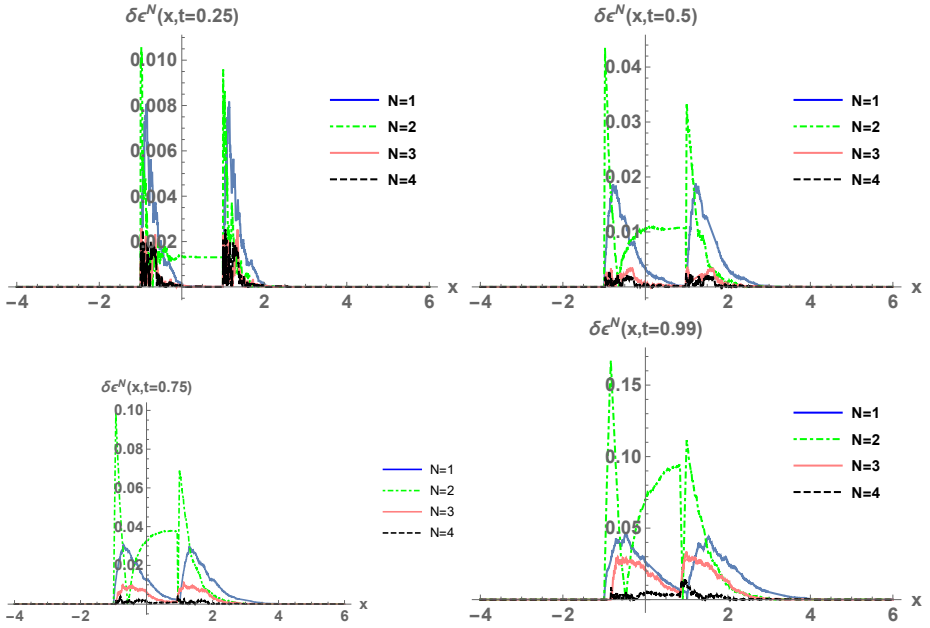

Figure 6.15: Differences in consecutive estimates $\delta\epsilon^N(x,t)$ (see (6.7)) at $t = 0.25$, $t = 0.5$, $t = 0.75$ and $t = 0.99$, with orders of truncation $N = 1$–4 in all three cases. This figure corresponds to Example 6.16.

A richer analysis of the convergence in this example is provided in Figure 6.15 and Table 6.4, which depict the consecutive differences $\delta\epsilon^N(x,t)$ (see (6.7)) and their norm $\Delta\epsilon^N(t)$ (see (6.8)), respectively. Even though the errors are not decreasing monotonically to 0 (as in the previous Example 6.15), the convergence is evident and follows from Theorem 6.11.

Finally, in Figure 6.16 we plot the error estimate $\log E^N(t)$ (see (6.9)), for all the times considered. Similar to Example 6.14, the plot shows that this example needs small orders N for all $t \in (-1,1)$ because of the decay of $|t|^N$.

Example 6.17 In this final example, we deal with discrete uncertainties, under the setting of Theorem 6.12. Consider (6.10) with $A_0 = 4$, $A_1 = 2$, $B_0 = 0$, $B_1 = -1$, $Y_0 \sim$ Bernoulli(0.4) and $Y_1 \sim$ Uniform$(-1,1)$, being all independent.

By Theorem 3.4, there is a unique mean square solution $X(t) = \sum_{n=0}^{\infty} X_n t^n$ on $\mathbb{R}$. For each $t \neq 0$, the random variable $X(t)$ is absolutely continuous, due to the absolute continuity of Y_1. Theorem 6.12, with Y_1 playing the role of Y_0

	$N = 1$	$N = 2$	$N = 3$	$N = 4$
$t = 0.25$	0.00666810	0.00482414	0.00169730	0.00275401
$t = 0.5$	0.0266447	0.0301975	0.00469989	0.00524461
$t = 0.75$	0.0582143	0.0920184	0.0176119	0.00756466
$t = 0.99$	0.0947395	0.190933	0.0597380	0.0107980

Table 6.4: Norm $\Delta\epsilon^N(t)$ of differences in consecutive estimates (see (6.8)) at different times t and for different truncation orders N. This table corresponds to Example 6.16.

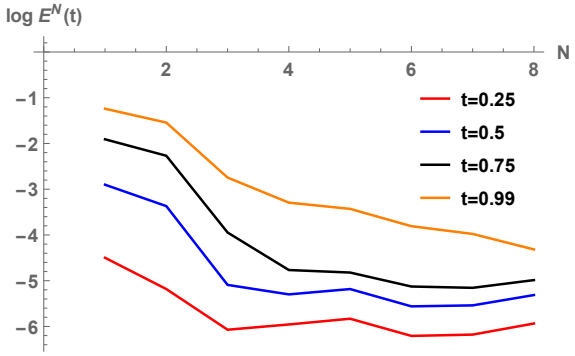

Figure 6.16: Error $E^N(t)$ (see (6.9)), for different times as indicated. This figure corresponds to Example 6.16.

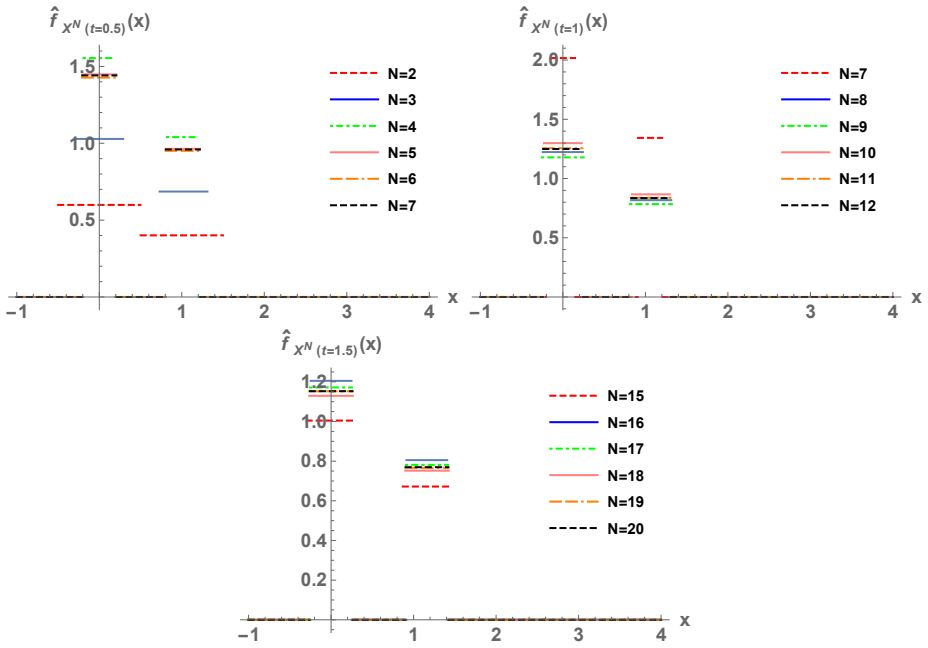

Figure 6.17: Graphical representations of the Monte Carlo estimates $\hat{f}_{X^N(t)}(x)$ at $t = 0.5$, $t = 1$ and $t = 1.5$, with orders of truncation N as indicated. This figure corresponds to Example 6.17.

(see Remark 6.10) allows approximating $f_{X(t)}(x)$ by utilizing the convergence $\lim_{N\to\infty} f_{X^N(t)}(x) = f_{X(t)}(x)$, which holds for almost every $x \in \mathbb{R}$.

In this particular example, Algorithm 1 is used with $M = 1,000,000$ iterations, because the deterministic values for $A(t)$ and $B(t)$ make the computational load much less demanding (see the discussion of Section 6.4). We will thus identify $\hat{f}_{X^N(t)}(x) = f_X^{N,M=1,000,000}(x,t)$.

For the times $t = 0.5$, 1 and 1.5, the numerical estimates $\hat{f}_{X^N(t)}(x)$ are displayed in Figure 6.17. Observe that the computed density functions are completely different to those of the previous examples: they are discontinuous, in fact step functions, mainly due to the discontinuities in f_{Y_1}. This example highlights the ability of our method to capture discontinuities.

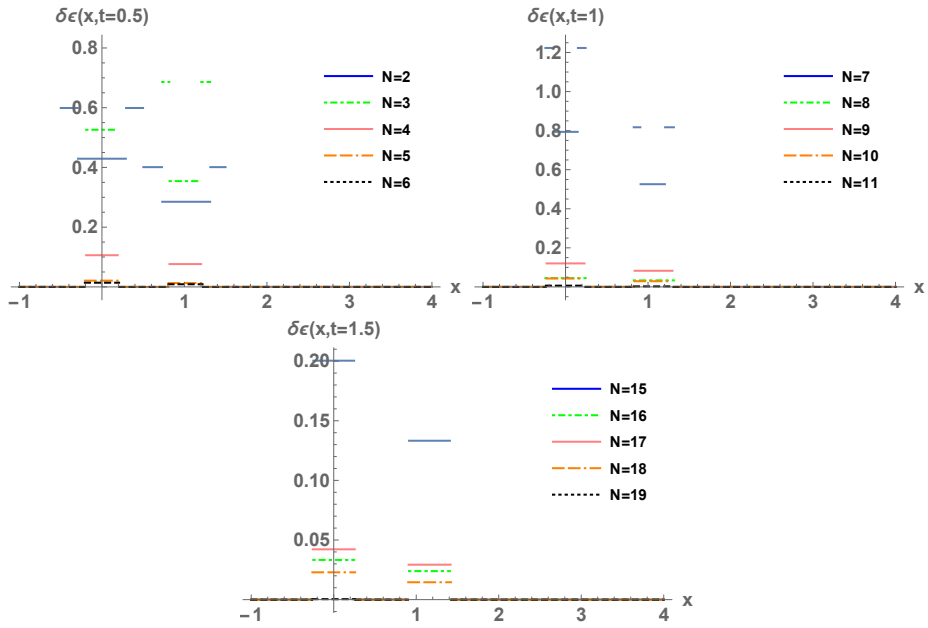

Figure 6.18: Differences in consecutive estimates $\delta\epsilon^N(x,t)$ (see (6.7)) at $t = 0.5$, $t = 1$ and $t = 1.5$, with orders of truncation N as indicated. This figure corresponds to Example 6.17.

The analysis of the convergence is completed with Figure 6.18 and Table 6.5, where the consecutive differences $\delta\epsilon^N(x,t)$ (see (6.7)) and their norms $\Delta\epsilon^N(t)$ (see (6.8)) are reported, respectively.

Finally, Figure 6.19 plots $\log E^N(t)$ (see (6.9)) as a function of N. The lower bound for the global error is the sampling error, which is smaller than in the previous four examples, owing to the larger number of samples considered. Comparing Figure 6.19 with the corresponding figures from the previous four examples, the non-monotonic decay of the error is also more pronounced for the three times. The discontinuity in the graph of the target distribution $f_{X(t)}(x)$ can explain the highly non-monotonic decay with the truncation order. Moreover, f_{Y_1} is not Lipschitz continuous, so the exponential convergence rate discussed in Remark 6.8 is not applicable in the present example. Finally, as for the other examples, the truncation order needed to reduce the error to the sampling contribution increases as we move away from the origin $t_0 = 0$. This behavior may pose severe challenges for large times t.

	$N = 2$	$N = 3$	$N = 4$	$N = 5$	$N = 6$
$t = 0.5$	0.833333	0.678571	0.140704	0.0280760	0.0190810
	$N = 7$	$N = 8$	$N = 9$	$N = 10$	$N = 11$
$t = 1$	0.785425	0.0767401	0.187360	0.0665995	0.00866170
	$N = 15$	$N = 16$	$N = 17$	$N = 18$	$N = 19$
$t = 1.5$	0.331941	0.0569325	0.0732120	0.0390644	0.00346696

Table 6.5: Norm $\Delta\epsilon^N(t)$ of differences in consecutive estimates (see (6.8)) at different times t and truncation orders N. This table corresponds to Example 6.17.

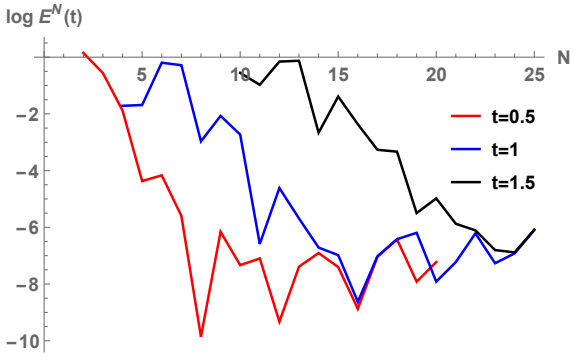

Figure 6.19: Error $E^N(t)$ (see (6.9)), for different times as indicated. This figure corresponds to Example 6.17.

6.6 Conclusions and perspectives

In this chapter, we address the analysis of the random non-autonomous second-order linear differential equation. When the data $A(t)$ and $B(t)$ are given by random power series on $(t_0 - r, t_0 + r)$ in $\mathrm{L}^\infty(\Omega)$, say $A(t) = \sum_{n=0}^{\infty} A_n (t - t_0)^n$ and $B(t) = \sum_{n=0}^{\infty} B_n (t - t_0)^n$, and the initial conditions Y_0 and Y_1 belong to $\mathrm{L}^2(\Omega)$, it is possible to construct a random power series solution $X(t) = \sum_{n=0}^{\infty} X_n (t - t_0)^n$ on $(t_0 - r, t_0 + r)$ in the mean square sense, whose coefficients satisfy a random difference equation. This approach is the generalization of the Fröbenius method to the random setting. The convergence rate of the power series of $X(t)$ is exponential for each time t, but not uniformly on the whole time domain $(t_0 - r, t_0 + r)$. For a fixed tolerance on the root mean square error of $X(t)$, the order of truncation of the power series needs to be increased, in general, when $|t - t_0|$ grows.

The probability density function of $X(t)$ can be expressed as the expectation of a random process $Z(x, t)$, $f_{X(t)}(x) = \mathbb{E}[Z(x, t)]$, using the law of total probability. A closed-form expression for Z is derived in terms of the fundamental set by exploiting the linearity of the system. However, to compute this expectation, one needs to perform a dimension reduction of the problem, by truncating the series of the fundamental set used to express the solution $X(t)$. Denoting $X^N(t)$, $N \geq 0$, the truncation of $X(t)$, we show that $f_{X^N(t)}(x)$ converges to $f_{X(t)}(x)$ pointwise as $N \to \infty$ under certain conditions (regarding Nemytskii operators); in some cases, an exponential convergence may be achieved for each t and x. The pointwise convergence also implies convergence in $\mathrm{L}^p(\mathbb{R})$, $1 \leq p < \infty$. In particular, the convergence in $\mathrm{L}^1(\mathbb{R})$ is equivalent to the convergence in the total variation and the Hellinger distances, which are instances of f-divergences.

From a numerical standpoint, the expectation defining $f_{X^N(t)}(x) = \mathbb{E}[Z^N(x, t)]$ is computable via a classical Monte Carlo strategy. We propose an algorithm for that purpose, which estimates symbolically $f_{X^N(t)}(x)$. This algorithm is implemented in the software Mathematica®, and it can be used to compute pointwise approximations of the density function $f_{X(t)}(x)$. One key feature of the algorithm is that it handles discontinuity and non-differentiability points of $f_{X(t)}(x)$ appropriately, without smoothing them out.

To the best of our knowledge, this work is the first one to provide such analysis of random second-order linear differential equations. However, we point out certain limitations of our methodology, which constitute potential avenues for future developments.

To start with, despite the exponential convergence rate, the approximations substantiated on the Fröbenius method may deteriorate for large $|t-t_0|$. This fact is inherent to Taylor series-based methods and also plagues other types of stochastic computations, such as PC expansions. Following [111], using random time-transformations may help to improve the convergence of the Fröbenius method and mitigate this issue.

Another point requiring a more in-depth analysis is the ignorance of the specific values of δ and μ. In particular, we showed that if the truncated processes from the fundamental set vanish for some trajectories near the time t of interest, the numerical estimate of the density becomes very noisy (see Example 6.15). This effect is due to the variance of $Z^N(x,t)$ that may be very large or infinite, with a severely deteriorated Monte Carlo convergence in these situations. In the following chapter we will explore different strategies to sort out this issue, such as the path-wise selection of the variable (Y_0 or Y_1) used in the expression of $Z^N(x,t)$, in order to control its variance.

Efforts to weaken or modify the theoretical hypotheses and enlarge the applicability of our method shall also be carried out. As an example, the extension of the method to the case of Y_0 and Y_1 not absolutely continuous would also present a valuable achievement. Similarly, an extension of the present methodology to linear systems of second-order random differential equations may be of great interest, while the application to other stochastic models of our expertise on random expansions and density approximations could be interesting.

At the computational level, the Monte Carlo estimation of $f_{X^N(t)}(x)$ introduces a statistical error since, in numerical computations, we are restricted to a finite number M of realizations. Therefore, an error of order $1/\sqrt{M}$ is unavoidable, even for $N \approx \infty$. The results presented in the chapter have highlighted the crucial importance of bias and sampling errors. It would be beneficial to rely on improved sampling strategies, such as multilevel Monte Carlo [75, 76], to balance the bias and sampling errors, while reducing the computational cost of the Monte Carlo estimates of the density. This topic is the focus of the following chapter.

Acknowledgements

This work has been supported by the Spanish "Ministerio de Economía y Competitividad" grant MTM2017–89664–P.

I acknowledge the doctorate scholarship granted by "Programa de Ayudas de Investigación y Desarrollo" (PAID), as well as "Ayudas para movilidad de estudiantes de doctorado de la Universitat Politècnica de València para estancias en 2019", Universitat Politècnica de València, for financing my research stay of three months at Centre de Mathématiques Appliquées (CMAP UMR7641), École Polytechnique, Palaiseau, France.

I am also grateful to Inria (Centre de Saclay and DeFi Team), which hosted me during my research stay at CMAP.

The main results of this chapter have been published in [94]. This chapter is an international collaboration with the researcher Olivier P. Le Maître (CMAP, CNRS, INRIA, École Polytechnique, Institut Polytechnique de Paris, 91128 Palaiseau, France).

Chapter 7

Variance reduction methods for Monte Carlo simulation in the density estimation of random second-order linear differential equations

This chapter concerns the estimation of the density function of the solution to a random non-autonomous second-order linear differential equation with analytic data processes. In the previous chapter, we proposed to express the density function as an expectation, and we used a standard Monte Carlo algorithm to approximate the expectation. Although the algorithms worked satisfactorily for most test-problems, some numerical challenges emerged for others, due to large statistical errors. In these situations, the convergence of the Monte Carlo simulation slows down severely, and noisy features plague the estimates. In this chapter, we focus on computational aspects and propose several variance reduction methods to remedy these issues and speed up the convergence. First, we introduce a pathwise selection of the approximating processes which aims at controlling the variance of the estimator. Second, we propose a hybrid method, combining Monte Carlo and deterministic quadrature rules, to estimate the expectation. Third, we exploit the series expansions of the solutions to design a multilevel Monte Carlo estimator. The proposed methods are implemented and tested on several numerical examples to highlight the theoretical discussions and demonstrate the significant improvements achieved.

7.1 Introduction

In Chapter 6, the expectation from (6.2) and (6.5) is approximated using Monte Carlo simulation, via an explicit algorithm. Either (6.2) or (6.5) are selected, and realizations of the involved random variables are generated to compute the sample average. This introduces a statistical error due to sampling, apart from the bias error $\theta_N(x,t) = f_{X(t)}(x) - f_{X^N(t)}(x)$. The convergence rate of the Monte Carlo procedure towards $f_{X^N(t)}(x)$ depends on the finite number of realizations and on the variances

$$\mathbb{V}\left[f_{Y_0}\left(\frac{x - Y_1 S_1^N(t)}{S_0^N(t)}\right)\frac{1}{|S_0^N(t)|}\right], \tag{7.1}$$

$$\mathbb{V}\left[f_{Y_1}\left(\frac{x - Y_0 S_0^N(t)}{S_1^N(t)}\right)\frac{1}{|S_1^N(t)|}\right]. \tag{7.2}$$

If these variances are large or infinite, the convergence rate of the Monte Carlo simulation deteriorates severely and noisy estimates of $f_{X^N(t)}(x)$ are produced, thus invalidating the results. This phenomenon was observed in the numerical experiments from Chapter 6. It is highly related to having small denominators $|S_0^N(t)|$ and $|S_1^N(t)|$ at certain realizable paths, as this may produce higher dispersion of $1/|S_0^N(t)|$ and $1/|S_1^N(t)|$.

The main goal of this chapter is to improve the algorithm from Chapter 6 via variance reduction methods to achieve good convergence of the Monte Carlo simulation and avoid noisy features. We will perform theoretical and numerical analysis of the new methods and algorithms, which will successfully enhance Chapter 6. The implementations and computations will be performed in the software Mathematica®, version 11.2.

In Section 7.2, we will path-wise select the initial condition used in the density expressions (6.2), (6.5), to avoid the smallest denominator between $|S_0^N(t)|$ and $|S_1^N(t)|$. In this manner, the variance of the statistical error will be controlled. We will thus be able to improve the quality of the numerical results from Chapter 6 significantly. An important assumption to apply this selection is that both initial conditions have probability density functions.

In Section 7.3, we will combine the algorithm from Section 7.2 with a quadrature rule for integration to compute the expectation with respect to Y_0–Y_1. We will see that this allows for reduction of variance, especially when Y_0–Y_1 are responsible for most of the estimator variability (this can be checked using Sobol indices). A detailed comparison with the methodology from Section 7.2, with regard to variance reduction, density approximation and error versus complex-

ity, will be carried out. In this chapter, complexity = cost = work, to avoid any confusion with the wording.

In Section 7.4, the Monte Carlo approach from Section 7.2 will be developed with a multilevel strategy. Originally introduced in [75, 76], the idea of multilevel Monte Carlo is to accelerate the convergence of the Monte Carlo simulation by decomposing the target expectation into the sum of expectations of increments (through a telescopic sum identity) whose variances decay rapidly. Multilevel Monte Carlo simulation is an improved sampling strategy that balances the bias and statistical errors, to obtain an overall reduction of the computational complexity compared to the standard Monte Carlo algorithm. Multilevel sampling strategies have been employed for direct approximation of statistics in stochastic differential equations and random partial differential equations, where the bias classically arises from a mesh discretization [8, 42, 69, 131]. However, to our knowledge, the application to the density estimation of random differential equations solution is completely original.

Finally, Section 7.5 will draw the main conclusions of the chapter. The methodologies and the algorithms proposed in the present chapter will be discussed. Potential avenues for future research on algorithmic improvements, linear random differential equations and density approximations will be suggested.

7.2 Variance reduction by path-wise selection of the initial condition used in the density expression

Consider the random processes

$$Z_0(x,t) = f_{Y_0}\left(\frac{x - Y_1 S_1(t)}{S_0(t)}\right)\frac{1}{|S_0(t)|},\ Z_1(x,t) = f_{Y_1}\left(\frac{x - Y_0 S_0(t)}{S_1(t)}\right)\frac{1}{|S_1(t)|}, \tag{7.3}$$

and

$$Z_0^N(x,t) = f_{Y_0}\left(\frac{x - Y_1 S_1^N(t)}{S_0^N(t)}\right)\frac{1}{|S_0^N(t)|},\ Z_1^N(x,t) = f_{Y_1}\left(\frac{x - Y_0 S_0^N(t)}{S_1^N(t)}\right)\frac{1}{|S_1^N(t)|}, \tag{7.4}$$

for $N \geq 0$. Here we are assuming that both initial conditions Y_0 and Y_1 have probability density functions f_{Y_0} and f_{Y_1}. The probability density function of $X(t)$ can be expressed through the expectation of $Z_0(x,t)$ or $Z_1(x,t)$, if $S_0(t) \neq 0$ or $S_1(t) \neq 0$ almost surely, respectively (see (6.1) and (6.4)):

$$f_{X(t)}(x) = \mathbb{E}[Z_0(x,t)] = \mathbb{E}[Z_1(x,t)].$$

When the power series $X(t)$ gets truncated, we obtain $X^N(t)$, whose probability density function is expressed through the expectation of $Z_0^N(x,t)$ or $Z_1^N(x,t)$, assuming that $S_0^N(t) \neq 0$ or $S_1^N(t) \neq 0$ almost surely, respectively (see (6.2) and (6.5)):

$$f_{X^N(t)}(x) = \mathbb{E}[Z_0^N(x,t)] = \mathbb{E}[Z_1^N(x,t)].$$

As justified in the previous chapter, $f_{X^N(t)}(x)$ tends to $f_{X(t)}(x)$ as $N \to \infty$, pointwise and in $\mathrm{L}^p(\mathbb{R}, \mathrm{d}x)$, $1 \leq p < \infty$. Under certain conditions (for instance, if f_{Y_0} or f_{Y_1} are Hölder continuous on $\mathbb{R}$), the pointwise convergence turns out to be exponentially fast, although not uniform in t and x.

In Chapter 6, the density function $f_{X^N(t)}(x)$ is approximated using a standard Monte Carlo sampling strategy, by initially selecting either $Z_0^N(x,t)$ or $Z_1^N(x,t)$ (only one of these processes). The convergence rate of the sample average towards $\mathbb{E}[Z_0^N(x,t)]$ or $\mathbb{E}[Z_1^N(x,t)]$ depends on $\mathbb{V}[Z_0^N(x,t)]$ or $\mathbb{V}[Z_1^N(x,t)]$ (see (7.1) and (7.2)). If $\mathbb{V}[Z_0^N(x,t)]$ and $\mathbb{V}[Z_1^N(x,t)]$ are large or infinity, the convergence rate deteriorates severely and noisy features may plague the statistical approximation to $f_{X^N(t)}(x)$. This issue was observed in the previous chapter.

In this section we propose a variance reduction method by combining the choices of $Z_0^N(x,t)$ and $Z_1^N(x,t)$. Consider one realization of $S_0^N(t)$ and $S_1^N(t)$. These realizations are computed recursively in the computer from realizations of $A_0, \ldots, A_{N-2}$, $B_0, \ldots, B_{N-2}$, see Chapter 6. If $|S_0^N(t)| \geq |S_1^N(t)|$ then we pick $Z_0^N(x,t)$, and vice versa otherwise. In such a way, we avoid the smallest denominator in the expressions of $Z_0^N(x,t)$ and $Z_1^N(x,t)$.

We assume that Y_0, Y_1 and $(A_0, A_1, \ldots, B_0, B_1, \ldots)$ are independent. We also suppose that Y_0 and Y_1 are absolutely continuous random variables (otherwise this method is no longer applicable). Given the events

$$G_0(t) = \{|S_0(t)| \geq |S_1(t)|\}, \quad G_1(t) = \{|S_1(t)| > |S_0(t)|\},$$

$$G_0^N(t) = \{|S_0^N(t)| \geq |S_1^N(t)|\}, \quad G_1^N(t) = \{|S_1^N(t)| > |S_0^N(t)|\},$$

we define the random processes

$$Z(x,t) = Z_0(x,t)\mathbb{1}_{G_0(t)} + Z_1(x,t)\mathbb{1}_{G_1(t)}, \tag{7.5}$$

$$Z^N(x,t) = Z_0^N(x,t)\mathbb{1}_{G_0^N(t)} + Z_1^N(x,t)\mathbb{1}_{G_1^N(t)}. \tag{7.6}$$

These processes correspond to path-wise selecting f_{Y_0} or f_{Y_1} according to $|S_0(t)| \geq |S_1(t)|$ or $|S_1(t)| > |S_0(t)|$ (in (7.5)), and $|S_0^N(t)| \geq |S_1^N(t)|$ or

$|S_1^N(t)| > |S_0^N(t)|$ (in (7.6)). The density functions of $X^N(t)$ and $X(t)$ are expressed as:

$$f_{X^N(t)}(x) = \mathbb{E}\left[Z^N(x,t)\right] \approx f_{X(t)}(x) = \mathbb{E}\left[Z(x,t)\right]. \tag{7.7}$$

Let us formalize (7.7) with Lemma 7.1 and Proposition (7.2), by using conditional probabilities.

Lemma 7.1 *Let U be an absolutely continuous random variable, (Z_1, Z_2) be a random vector such that $Z_1 \neq 0$ almost surely, and $G \in \mathcal{F}$ be an event of non-zero probability. Suppose that U and $(Z_1, Z_2, \mathbb{1}_G)$ are independent. Then the probability law of $Z_1U + Z_2$ conditioned to G is absolutely continuous, with density function*

$$f_{Z_1U+Z_2}(z|G) = \mathbb{E}\left[f_U\left(\frac{z - Z_2}{Z_1}\right)\frac{1}{|Z_1|}\middle| G\right].$$

Proof. On the measurable space $(\Omega, \mathcal{F})$, define the conditional probability measure $\mathbb{P}_G$ as $\mathbb{P}_G[E] = \mathbb{P}[E|G] := \mathbb{P}[E \cap G]/\mathbb{P}[G]$, for $E \in \mathcal{F}$.
Since U and G are independent, the probability law of U with respect to $\mathbb{P}_G$ coincides with $\mathbb{P}$. Then the law of U under $\mathbb{P}_G$ is absolutely continuous, and $f_U(u|G) = f_U(u)$.
On the other hand, notice that U and (Z_1, Z_2) are independent under $\mathbb{P}_G$.
Finally, $\mathbb{P}_G$ is absolutely continuous with respect to $\mathbb{P}$, $\mathbb{P}_G \ll \mathbb{P}$, therefore $Z_1 \neq 0$ $\mathbb{P}_G$-almost surely.
Thus, on the probability space $(\Omega, \mathcal{F}, \mathbb{P}_G)$ we are in conditions of applying Lemma 6.2:

$$f_{Z_1U+Z_2}(z|G) = \mathbb{E}\left[f_U\left(\frac{z - Z_2}{Z_1}\middle| G\right)\frac{1}{|Z_1|}\middle| G\right] = \mathbb{E}\left[f_U\left(\frac{z - Z_2}{Z_1}\right)\frac{1}{|Z_1|}\middle| G\right].$$

□

Proposition 7.2 *The following relationships hold:*

$$f_{X(t)}(x) = \mathbb{E}\left[Z(x,t)\right], \quad f_{X^N(t)}(x) = \mathbb{E}\left[Z^N(x,t)\right],$$

where $Z(x,t)$ and $Z^N(x,t)$ are defined in (7.5) *and* (7.6).

Proof. We focus on the equality $f_{X(t)}(x) = \mathbb{E}[Z(x,t)]$, as the other one concerning $f_{X^N(t)}(x)$ is analogous.
If either $\mathbb{P}[G_0(t)] = 0$ or $\mathbb{P}[G_1(t)] = 0$, then $Z(x,t) = Z_1(x,t)$ almost surely or $Z(x,t) = Z_0(x,t)$ almost surely, respectively, so we are done.

Let us then assume that both $G_0(t)$ and $G_1(t)$ have non-zero probability. By the law of total probability,

$$f_{X(t)}(x) = f_{X(t)}\left(x|G_0(t)\right)\mathbb{P}[G_0(t)] + f_{X(t)}\left(x|G_1(t)\right)\mathbb{P}[G_1(t)].$$

To compute the conditional densities, we use Lemma 7.1:

$$f_{X(t)}\left(x|G_0(t)\right) = \mathbb{E}\left[Z_0(x,t)|G_0(t)\right],$$
$$f_{X(t)}\left(x|G_1(t)\right) = \mathbb{E}\left[Z_1(x,t)|G_1(t)\right].$$

As a consequence,

$$\begin{aligned} f_{X(t)}(x) &= \mathbb{E}\left[Z_0(x,t)|G_0(t)\right]\mathbb{P}[G_0(t)] + \mathbb{E}\left[Z_1(x,t)|G_1(t)\right]\mathbb{P}[G_1(t)] \\ &= \mathbb{E}\left[Z_0(x,t)\mathbb{1}_{G_0(t)}\right] + \mathbb{E}\left[Z_1(x,t)\mathbb{1}_{G_1(t)}\right] = \mathbb{E}\left[Z(x,t)\right]. \end{aligned}$$

□

The expectation $f_{X^N(t)}(x) = \mathbb{E}[Z^N(x,t)]$ can be approximated via a standard Monte Carlo sampling strategy. A key feature here is the uniform boundedness of $\mathbb{V}[Z^N(x,t)]$ with N, as shown in Proposition 7.4.

Lemma 7.3 *Given any $T > t_0$, there exists a constant $\alpha > 0$ such that*

$$\max\{|S_0(t)|, |S_1(t)|\} \geq 2\alpha \tag{7.8}$$

holds almost surely, for all $t \in [t_0, T]$. In particular, there exists an integer $N_{T,\alpha} \geq 0$ such that

$$\max\{|S_0^N(t)|, |S_1^N(t)|\} \geq \alpha \tag{7.9}$$

holds almost surely, for all $t \in [t_0, T]$ and $N \geq N_{T,\alpha}$.

Proof. Fix $T > t_0$ and $t \in [t_0, T]$. Consider the Wronskian $W_{S_0,S_1}(t) = S_0(t)\dot{S}_1(t) - S_1(t)\dot{S}_0(t)$, where the derivatives are regarded in the $\mathrm{L}^\infty(\Omega)$ sense. Recall that $S_0(t)$ and $S_1(t)$ solve the differential equation in the $\mathrm{L}^\infty(\Omega)$ sense, see Section 7.1. In particular, the sample paths of $S_0(t)$ and $S_1(t)$ solve the deterministic version of the differential equation, for almost every $\omega \in \Omega$. By Liouville's formula for deterministic linear differential equations [41, Prop. 2.15],

$$W_{S_0,S_1}(t) = \mathrm{e}^{-\int_{t_0}^t A(s)\,\mathrm{d}s}$$

almost surely. As $A(t)$ is a power series in $\mathrm{L}^\infty(\Omega)$, we can lower bound the Wronskian as follows:

$$W_{S_0,S_1}(t) \geq \mathrm{e}^{-\int_{t_0}^T \|A(s)\|_\infty\,\mathrm{d}s} =: \beta > 0,$$

almost surely. On the other hand, using the triangular inequality,

$$\begin{aligned} W_{S_0,S_1}(t) &\leq 2\max\{|S_0(t)|,|S_1(t)|\}\max\{|\dot{S}_0(t)|,|\dot{S}_1(t)|\} \\ &\leq 2\max\{|S_0(t)|,|S_1(t)|\}\max\{\|\dot{S}_0(t)\|_\infty,\|\dot{S}_1(t)\|_\infty\} \\ &\leq 2\max\{|S_0(t)|,|S_1(t)|\}\gamma, \end{aligned}$$

where γ is a constant upper bound for $\|\dot{S}_0(t)\|_\infty$ and $\|\dot{S}_1(t)\|_\infty$ on $[t_0,T]$. As a consequence,

$$\max\{|S_0(t)|,|S_1(t)|\} \geq \frac{\beta}{2\gamma} =: 2\alpha,$$

almost surely, and (7.8) is proved.
Finally, we have that $\max\{|S_0^N(t)|,|S_1^N(t)|\}$ converges to $\max\{|S_0(t)|,|S_1(t)|\}$ as $N\to\infty$ in $\mathrm{L}^\infty(\Omega)$, uniformly on $[t_0,T]$. This fact allows deducing that there exists an integer $N_{T,\alpha}\geq 0$ such that (7.9) holds.

□

Proposition 7.4 *Suppose that f_{Y_0} and f_{Y_1} are bounded density functions. Given any $T>t_0$, the $\mathrm{L}^2(\Omega)$ norms of $Z(x,t)$ and $Z^N(x,t)$ are controlled as follows:*

$$\|Z(x,t)\|_2 \leq \frac{\max\{\|f_{Y_0}\|_\infty,\|f_{Y_1}\|_\infty\}}{2\alpha},$$
$$\|Z^N(x,t)\|_2 \leq \frac{\max\{\|f_{Y_0}\|_\infty,\|f_{Y_1}\|_\infty\}}{\alpha},$$

for $x\in\mathbb{R}$, $t\in[t_0,T]$ and $N\geq N_{T,\alpha}$, where α is the constant from Lemma 7.3.

Proof. Observe that, by (7.5) and (7.6),

$$\begin{aligned} |Z(x,t)| &\leq \max\{\|f_{Y_0}\|_\infty,\|f_{Y_1}\|_\infty\}\left(\frac{1}{|S_0(t)|}\mathbb{1}_{G_0(t)}+\frac{1}{|S_1(t)|}\mathbb{1}_{G_1(t)}\right) \\ &= \frac{\max\{\|f_{Y_0}\|_\infty,\|f_{Y_1}\|_\infty\}}{\max\{|S_0(t)|,|S_1(t)|\}}, \end{aligned}$$

$$\begin{aligned} |Z^N(x,t)| &\leq \max\{\|f_{Y_0}\|_\infty,\|f_{Y_1}\|_\infty\}\left(\frac{1}{|S_0^N(t)|}\mathbb{1}_{G_0^N(t)}+\frac{1}{|S_1^N(t)|}\mathbb{1}_{G_1^N(t)}\right) \\ &= \frac{\max\{\|f_{Y_0}\|_\infty,\|f_{Y_1}\|_\infty\}}{\max\{|S_0^N(t)|,|S_1^N(t)|\}}. \end{aligned}$$

Then Lemma 7.3 applies.

□

As a consequence, the variance here is well controlled, in contrast with Chapter 6. The estimate of $f_{X^N(t)}(x) = \mathbb{E}[Z^N(x,t)]$ using Monte Carlo simulation does not present convergence problems.

Remark 7.5 *Lemma 7.3 is not only important from a numerical point of view, but also for the following theoretical fact. Chapter 6 justified the pointwise convergence of* $\{f_{X^N(t)}(x)\}_{N=0}^{\infty}$ *towards* $f_{X(t)}(x)$ *in a neighborhood of* t_0*. The neighborhood was constructed in such a way that* $S_0(t)$ *and/or* $S_1(t)$ *are greater than a positive constant almost surely (see Remarks 6.5 and 6.10 in Chapter 6), so that the denominators from* (7.4) *are controlled. Lemma 7.3 shows that*

$$\frac{1}{|S_0(t)|}\mathbb{1}_{G_0(t)} + \frac{1}{|S_1(t)|}\mathbb{1}_{G_1(t)} = \frac{1}{\max\{|S_0(t)|, |S_1(t)|\}}$$

is upper bounded by a finite constant almost surely, for every t *in the domain of definition of* (3.1)*. Hence, the theoretical results from Chapter 6 justifying the convergence in* (7.7) *as* $N \to \infty$ *do hold for every* t *where* $X(t)$ *is well-defined, not just in a neighborhood of* t_0*. Notice that this statement requires both* Y_0 *and* Y_1 *to be absolutely continuous, and* Y_0*,* Y_1 *and* $(A_0, A_1, \ldots, B_0, B_1, \ldots)$ *to be independent. These are the hypotheses under which the methodology from this section is supported.*

The whole procedure described in this section is structured in the form of Algorithm 2. By path-wise selecting $Z_0^N(x,t)$ or $Z_1^N(x,t)$ according to $|S_0^N(t)| \geq |S_1^N(t)|$ or $|S_1^N(t)| > |S_0^N(t)|$, we approximate the expectation $f_{X^N(t)}(x) = \mathbb{E}[Z^N(x,t)]$ using a sample average of M realizations. The output function is denoted as $f_X^{N,M}(x,t)$, which tends to $f_{X^N(t)}(x)$ as $M \to \infty$.

The complexity (cost, work) of Algorithm 2 is $\mathcal{O}(MN^2)$, in general. Recall that each realization of $S_0^N(t)$ and $S_1^N(t)$ is computed recursively with N and requires $\mathcal{O}(N^2)$ operations (see Chapter 6).

As discussed in Chapter 6, when $A(t)$ and $B(t)$ are random polynomials instead of infinite random series, the cost per realization of $S_0^N(t)$ and $S_1^N(t)$ is reduced to $\mathcal{O}(N)$ operations. Then the global complexity of Algorithm 2 is $\mathcal{O}(MN)$ only.

Obviously, in the case where $S_0(t)$ and $S_1(t)$ are known in closed-form expression (this occurs for simple problems (3.1), such as the autonomous case), Algorithm 2 can be run with $N = \infty$, that is, by computing realizations of $S_0(t)$ and $S_1(t)$ directly. The complexity of the algorithm is $\mathcal{O}(M)$ in such a case. The error is statistical and only due to the finite sampling.

Algorithm 2 Estimation of $f_{X^N(t)}(x)$ using Monte Carlo simulation, Section 7.2.

Inputs: t_0; N; f_{Y_0}; f_{Y_1}; probability distribution of $A_0, \ldots, A_N$, $B_0, \ldots, B_N$; number M of realizations in the Monte Carlo procedure; t; and discretization vector x of the density domain.

$\Sigma \leftarrow 0$ ▷ Initialize the samples sum
for $i = 1, \ldots, M$ **do** ▷ Monte Carlo loop
  Set a realization of $S_0^N(t)$ and $S_1^N(t)$
  if $|S_0^N(t)| \geq |S_1^N(t)|$ **then**
    Set a realization of Y_1
    $\Sigma \leftarrow \Sigma + Z_0^N(x,t)$ ▷ See (7.4)
  else
    Set a realization of Y_0
    $\Sigma \leftarrow \Sigma + Z_1^N(x,t)$ ▷ See (7.4)
  end if
end for
$f_X^{N,M}(x,t) \leftarrow \Sigma/M$ ▷ Set sample average
Return $f_X^{N,M}(x,t)$ ▷ Approximation of $f_{X^N(t)}(x)$

Algorithm 2 may be run symbolically on the variable x, thus having a symbolic expression for the output $f_X^{N,M}(x,t)$ in the end. The algorithm from Chapter 6 was run symbolically. However, symbolic computations do not allow a large number of simulations M. Thus, in cases where M needs to be large enough to achieve convergence, we recommend to discretize x in a domain of interest for the target density.

In the following examples, we illustrate the theoretical discussion from this section. The main goal is to highlight the improvement of Algorithm 2 compared with Chapter 6.

Example 7.6 We start by considering a very simple problem (3.1):

$$\begin{cases} \ddot{X}(t) + B^2 X(t) = 0, \ t \in \mathbb{R}, \\ X(t_0 = 0) = Y_0, \\ \dot{X}(t_0 = 0) = Y_1. \end{cases} \tag{7.10}$$

Despite its simplicity, (7.10) is a useful example to illustrate the theoretical discussion from this section.

The following probability distributions are taken for the random inputs: $B \sim \text{Uniform}(1, 2.5)$, $Y_0, Y_1 \sim \text{Normal}(1, \sigma = 0.1)$, all of them independent. As

B is bounded and the initial conditions Y_0 and Y_1 have absolute moments of any order, there exists an analytic solution $X(t)$ in the $\mathrm{L}^p(\Omega)$ sense, for $1 \leq p < \infty$. The fundamental system is explicitly known: $S_0(t) = \cos(Bt)$, $S_1(t) = \sin(Bt)/B$, $t \in \mathbb{R}$. Thus, Algorithm 2 can be applied with $N = \infty$, by taking realizations from $S_0(t)$ and $S_1(t)$ directly. We consider time $t = 10$, and the goal is to approximate the density function $f_{X(t=10)}(x)$ using Monte Carlo methods.

In Figure 7.1, we estimate $\mathbb{E}[Z_0(x, t = 10)]$ and $\mathbb{E}[Z_1(x, t = 10)]$ with the Monte Carlo procedure from Chapter 6, by generating $M = 100,000$ realizations of the random inputs. These expectations correspond to $f_{X(t=10)}(x)$. Observe that noisy features plague the estimates of the density function. In Figure 7.2, we plot estimates of the variances $\mathbb{V}[Z_0(x, t = 10)]$ and $\mathbb{V}[Z_1(x, t = 10)]$. Their large values slow down the convergence of the Monte Carlo simulation, thus explaining the noisy behavior of the estimates in Figure 7.1. In fact, at the points x where the variance is larger, more noise is perceived in the estimate of the expectation.

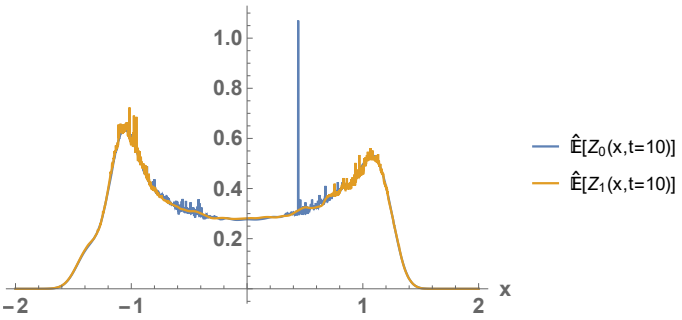

Figure 7.1: Approximations of $f_{X(t=10)}(x)$ using estimates $\hat{\mathbb{E}}[Z_0(x, t = 10)]$ and $\hat{\mathbb{E}}[Z_1(x, t = 10)]$. This figure corresponds to Example 7.6.

In Figure 7.3, we approximate $f_{X(t=10)}(x)$ via $\mathbb{E}[Z(x, t = 10)]$, by using the proposed Algorithm 2 with $M = 100,000$ realizations. Observe that now the estimate is smooth, thus resolving the convergence challenge. In Figure 7.4, we depict an estimate for $\mathbb{V}[Z(x, t = 10)]$. This variance is shown to be small, thus allowing for faster convergence of the Monte Carlo simulation. Hence, Algorithm 2 improves Chapter 6 significantly.

It is interesting to observe the different behaviors of $Z_0(x, t = 10)$, $Z_1(x, t = 10)$ and $Z(x, t = 10) = Z_0(x, t = 10)\mathbb{1}_{G_0(t=10)} + Z_1(x, t = 10)\mathbb{1}_{G_1(t=10)}$ as real functions of B, Y_0 and Y_1. Figure 7.5 depicts the 3D graphs of $Z_0(x =$

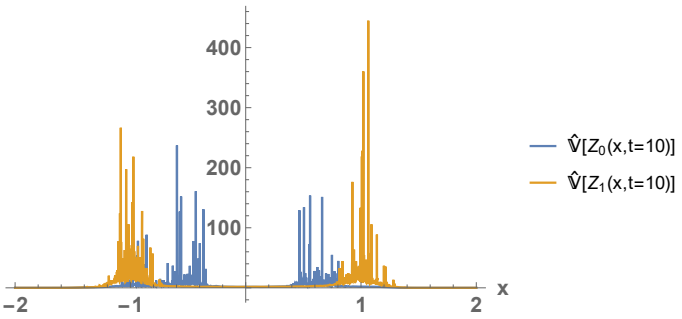

Figure 7.2: Estimates $\hat{\mathbb{V}}[Z_0(x, t = 10)]$ and $\hat{\mathbb{V}}[Z_1(x, t = 10)]$. This figure corresponds to Example 7.6.

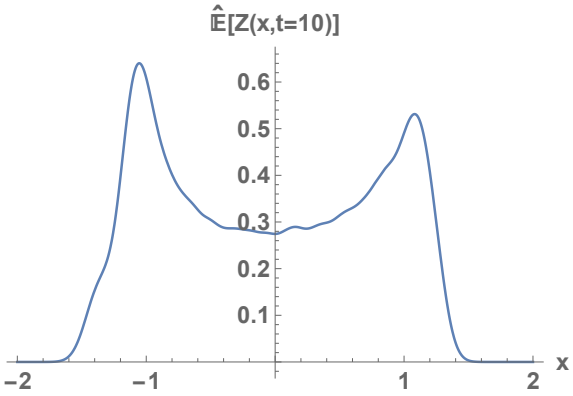

Figure 7.3: Approximation of $f_{X(t=10)}(x)$ using estimate $\hat{\mathbb{E}}[Z(x, t = 10)]$. This figure corresponds to Example 7.6.

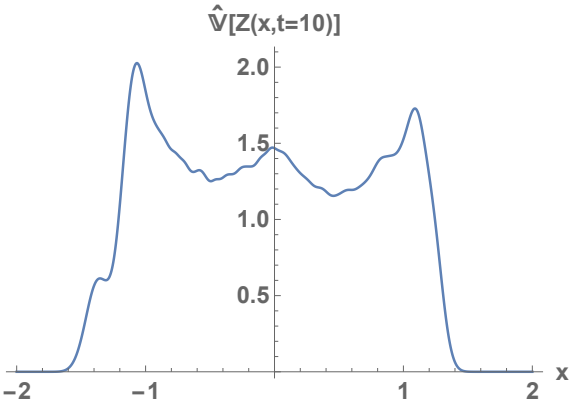

Figure 7.4: Estimate $\hat{\mathbb{V}}[Z(x, t = 10)]$. This figure corresponds to Example 7.6.

$0, t = 10)$, $Z_1(x = 0, t = 10)$, $Z_0(x = 0, t = 10)\mathbb{1}_{G_0(t=10)}$ and $Z_1(x = 0, t = 10)\mathbb{1}_{G_1(t=10)}$ (obviously, any other point x different from 0 can also be taken). The functions $Z_0(x = 0, t = 10)$ and $Z_1(x = 0, t = 10)$ have singularities (because of vanishing denominators $|\cos(10B)|$ and $|\sin(10B)/B|$) and are strongly peaked. On the contrary, although $Z_0(x = 0, t = 10)\mathbb{1}_{G_0(t=10)}$ and $Z_1(x = 0, t = 10)\mathbb{1}_{G_1(t=10)}$ possess discontinuity points, their peaks are not that sharp. As we are getting rid of small denominators, the sharp peaks are avoided.

Example 7.7 We consider a more complex example now, based on degree one polynomial coefficients. Let

$$\begin{cases} \ddot{X}(t) + (A_0 + A_1 t)\dot{X}(t) + (B_0 + B_1 t)X(t) = 0, \ t \in \mathbb{R}, \\ X(t_0 = 0) = Y_0, \\ \dot{X}(t_0 = 0) = Y_1, \end{cases} \tag{7.11}$$

$A_0 = 4$, $A_1 \sim \text{Uniform}(0, 1)$, $B_0 \sim \text{Gamma}(2, 2)|_{[0,4]}$, $B_1 \sim \text{Bernoulli}(0.35)$ and $Y_0, Y_1 \sim \text{Normal}(2, 1)$, all of them independent. Since the random coefficients A_0, A_1, B_0 and B_1 are bounded and the initial conditions Y_0 and Y_1 have finite absolute moments, there is an analytic solution $X(t)$ in the $\mathrm{L}^p(\Omega)$ sense, $1 \leq p < \infty$.

According to Chapter 6, the probability density function $f_{X(t)}(x)$ can be approximated pointwise using $f_{X^N(t)}(x)$ (with exponential convergence with N because f_{Y_0} and f_{Y_1} are Lipschitz continuous on $\mathbb{R}$). We focus on the order

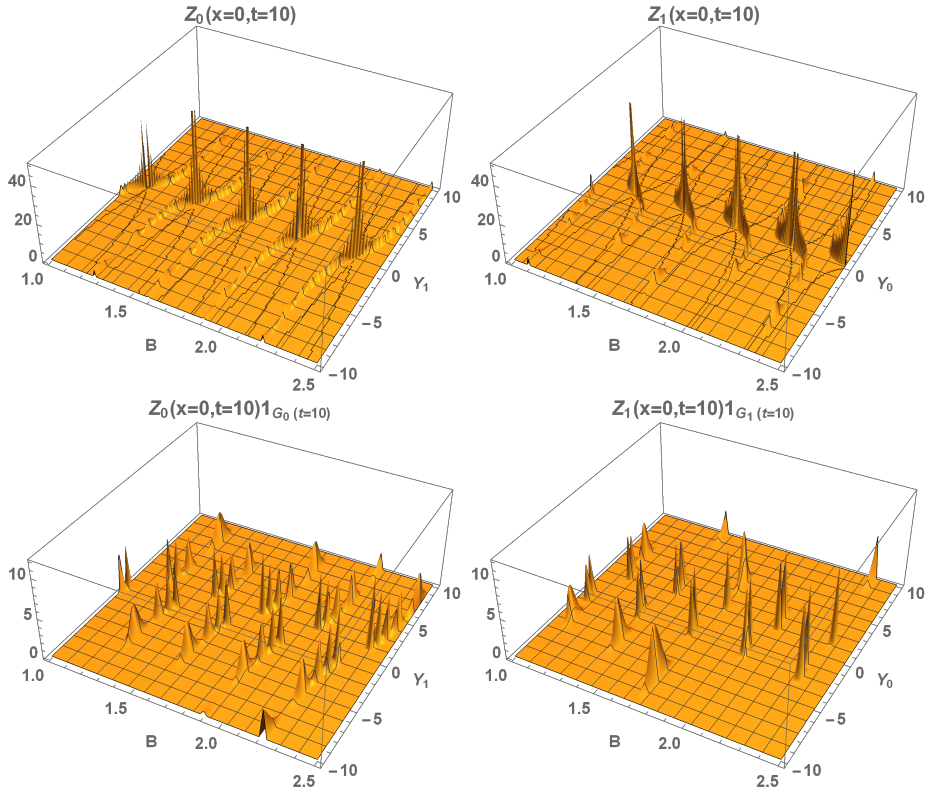

Figure 7.5: 3D plots of the integrands $Z_0(x = 0, t = 10)$, $Z_1(x = 0, t = 10)$ (their vertical axes have been restricted to $[0, 50]$), $Z_0(x = 0, t = 10)\mathbb{1}_{G_0(t=10)}$ and $Z_1(x = 0, t = 10)\mathbb{1}_{G_1(t=10)}$. This figure corresponds to Example 7.6.

of truncation $N = 10$ to highlight the improvement of Algorithm 2 compared with Chapter 6. We estimate $f_{X^{N=10}(t=1.5)}(x)$ with both approaches.

In Figure 7.6 we estimate $\mathbb{E}[Z_0^{N=10}(x, t = 1.5)]$ and $\mathbb{E}[Z_1^{N=10}(x, t = 1.5)]$ with the algorithm described in Chapter 6. With $M = 20,000$ realizations, we observe noise in both estimates. This phenomenon is accurately explained by the plots of $\mathbb{V}[Z_0^{N=10}(x, t = 1.5)]$ and $\mathbb{V}[Z_1^{N=10}(x, t = 1.5)]$ in Figure 7.7. It is very interesting to observe that the points x where the variances are larger correspond to higher noise in Figure 7.6. Indeed, large variance affects the rate of convergence of the Monte Carlo simulation severely.

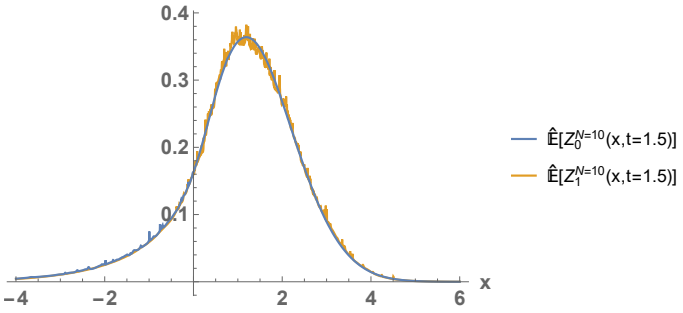

Figure 7.6: Approximations of $f_{X^{N=10}(t=1.5)}(x)$ using estimates $\hat{\mathbb{E}}[Z_0^{N=10}(x, t = 1.5)]$ and $\hat{\mathbb{E}}[Z_1^{N=10}(x, t = 1.5)]$. This figure corresponds to Example 7.7.

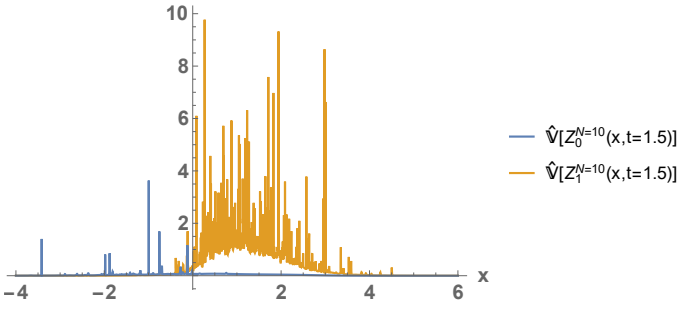

Figure 7.7: Estimates $\hat{\mathbb{V}}[Z_0^{N=10}(x, t = 1.5)]$ and $\hat{\mathbb{V}}[Z_1^{N=10}(x, t = 1.5)]$. This figure corresponds to Example 7.7.

Algorithm 2 has been designed to fix this issue. Figure 7.8 shows the smooth estimate of $f_{X^{N=10}(t=1.5)}(x)$ with $\mathbb{E}[Z^{N=10}(x, t = 1.5)]$ and $M = 20,000$ realizations. This good convergence is due to the small variance $\mathbb{V}[Z^{N=10}(x, t = 1.5)]$,

see Figure 7.9. Hence, Algorithm 2 shows significantly superior to a mere Monte Carlo approach as in Chapter 6.

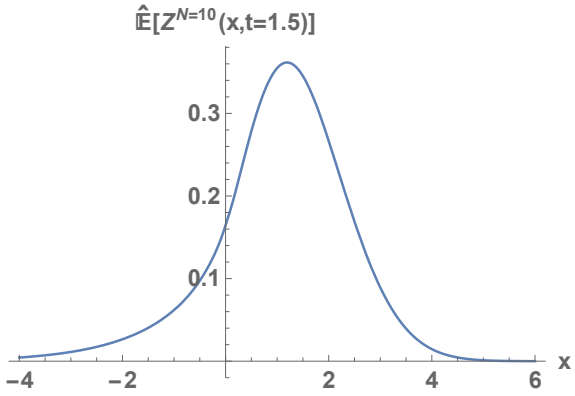

Figure 7.8: Approximation of $f_{X^{N=10}(t=1.5)}(x)$ using estimate $\hat{\mathbb{E}}[Z^{N=10}(x, t = 1.5)]$. This figure corresponds to Example 7.7.

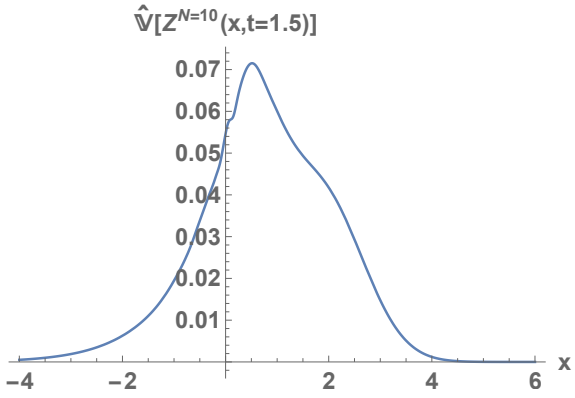

Figure 7.9: Estimate $\hat{\mathbb{V}}[Z^{N=10}(x, t = 1.5)]$. This figure corresponds to Example 7.7.

7.3 Variance reduction by using quadrature rule for the initial conditions

By Proposition 7.2, the probability density function of the truncated series $X^N(t)$ is expressed as $f_{X^N(t)}(x) = \mathbb{E}[Z^N(x,t)]$, where $Z^N(x,t)$ was defined in (7.6). The randomness in $Z^N(x,t)$ comes from the initial conditions, Y_0 and Y_1, and from the truncated series from the fundamental set, $S_0^N(t)$ and $S_1^N(t)$. A Monte Carlo sampling procedure generates M independent realizations from these random quantities and approximates $f_{X^N(t)}(x)$ using the sample average, see Algorithm 2.

The convergence rate of the sample mean towards the true expectation depends heavily on the variance of $Z^N(x,t)$ (which is uniformly bounded on N by Proposition 7.4). In this section, we present a method for variance reduction. Each time we generate a realization of $S_0^N(t)$ and $S_1^N(t)$ and check whether $|S_0^N(t)| \geq |S_1^N(t)|$ or $|S_1^N(t)| > |S_0^N(t)|$, now we do not generate a realization of Y_0–Y_1, but compute the expectation with respect to Y_0–Y_1 with a numerical integration method, instead.

We write $f_{X^N(t)}(x)$ as $f_{X^N(t)}(x) = \mathbb{E}[\mathbb{E}[Z^N(x,t)|S_0^N(t), S_1^N(t)]]$, by the law of total expectation. By a Monte Carlo sampling procedure, we approximate the outer expectation as

$$f_{X^N(t)}(x) \approx \frac{1}{M}\sum_{i=1}^{M} \mathbb{E}[Z^N(x,t)|S_0^N(t) = s_{0,i}^N, S_1^N(t) = s_{1,i}^N], \tag{7.12}$$

where $\{(s_{0,i}^N, s_{1,i}^N)\}_{i=1}^M$ are M independent realizations of $(S_0^N(t), S_1^N(t))$.

The variance of the statistical error from the approximation (7.12) is always smaller than in the standard Monte Carlo approach from Algorithm 2. Indeed, by the law of total variance,

$$\begin{aligned}\mathbb{V}\left[\mathbb{E}\left[Z^N(x,t)|S_0^N(t), S_1^N(t)\right]\right] &= \mathbb{V}\left[Z^N(x,t)\right] - \mathbb{E}\left[\mathbb{V}\left[Z^N(x,t)|S_0^N(t), S_1^N(t)\right]\right] \\ &\leq \mathbb{V}\left[Z^N(x,t)\right]. \end{aligned} \tag{7.13}$$

Equality holds only when $\mathbb{V}[Z^N(x,t)|S_0^N(t), S_1^N(t)] = 0$. Essentially, using (7.12) we gain more reduction in the statistical error when Y_0–Y_1 produce high variability in $Z^N(x,t)$. The extreme case occurs when $A(t)$ and $B(t)$ are deterministic functions, as in such a case a deterministic integration rule for Y_0–Y_1 makes the method sampling error-free.

The influence of the variability of Y_0, Y_1, $S_0^N(t)$ and $S_1^N(t)$ on $Z^N(x,t)$ (variance-based sensitivity analysis) can be assessed using Sobol indices [140, 158]. The

random process $Z^N = Z^N(x,t)$ has the following orthogonal decomposition:

$$Z^N = \mathbb{E}\left[Z^N\right] + \tilde{Z}^N_{S_0^N(t),S_1^N(t)} + \tilde{Z}^N_{Y_0,Y_1} + \tilde{Z}^N_{S_0^N(t),S_1^N(t),Y_0,Y_1},$$

where

$$\tilde{Z}^N_{S_0^N(t),S_1^N(t)} = Z^N_{S_0^N(t),S_1^N(t)} - \mathbb{E}\left[Z^N_{S_0^N(t),S_1^N(t)}\right],\ Z^N_{S_0^N(t),S_1^N(t)} = \mathbb{E}\left[Z^N | S_0^N(t), S_1^N(t)\right],$$

$$\tilde{Z}^N_{Y_0,Y_1} = Z^N_{Y_0,Y_1} - \mathbb{E}\left[Z^N_{Y_0,Y_1}\right],\ Z^N_{Y_0,Y_1} = \mathbb{E}\left[Z^N | Y_0, Y_1\right],$$

$$\tilde{Z}^N_{S_0^N(t),S_1^N(t),Y_0,Y_1} = Z^N_{S_0^N(t),S_1^N(t),Y_0,Y_1} - \mathbb{E}\left[Z^N_{S_0^N(t),S_1^N(t),Y_0,Y_1}\right],$$

$$Z^N_{S_0^N(t),S_1^N(t),Y_0,Y_1} = Z^N - \tilde{Z}^N_{S_0^N(t),S_1^N(t)} - \tilde{Z}^N_{Y_0,Y_1}.$$

Applying squared $\mathrm{L}^2(\Omega)$ norms, we derive the ANOVA-Hoeffding decomposition of the variance of $Z^N(x,t)$:

$$\mathbb{V}\left[Z^N\right] = \mathbb{V}\left[Z^N_{S_0^N(t),S_1^N(t)}\right] + \mathbb{V}\left[Z^N_{Y_0,Y_1}\right] + \mathbb{V}\left[Z^N_{S_0^N(t),S_1^N(t),Y_0,Y_1}\right],$$

where $\mathbb{V}[Z^N_{S_0^N(t),S_1^N(t)}]$, $\mathbb{V}[Z^N_{Y_0,Y_1}]$ and $\mathbb{V}[Z^N_{S_0^N(t),S_1^N(t),Y_0,Y_1}]$ measure the contribution to $Z^N(x,t)$ of $(S_0^N(t), S_1^N(t))$ alone, (Y_0, Y_1) alone, and the interactions between them, respectively. The standardization of the variance decomposition gives rise to the Sobol indices:

$$\mathbb{S}^N_{S_0^N(t),S_1^N(t)} = \frac{\mathbb{V}\left[Z^N_{S_0^N(t),S_1^N(t)}\right]}{\mathbb{V}\left[Z^N\right]}, \quad \mathbb{S}^N_{Y_0,Y_1} = \frac{\mathbb{V}\left[Z^N_{Y_0,Y_1}\right]}{\mathbb{V}\left[Z^N\right]}, \tag{7.14}$$

$$\mathbb{S}^N_{S_0^N(t),S_1^N(t),Y_0,Y_1} = \frac{\mathbb{V}\left[Z^N_{S_0^N(t),S_1^N(t),Y_0,Y_1}\right]}{\mathbb{V}\left[Z^N\right]}, \tag{7.15}$$

which are scaled in $[0,1]$. When $\mathbb{S}^N_{Y_0,Y_1}$ is near 1, we expect more variance reduction in the Monte Carlo estimate (7.12).

Fix the index of the realizations $i = 1, \ldots, M$. If $|s^N_{0,i}| \geq |s^N_{1,i}|$, then

$$\begin{aligned}\mathbb{E}\left[Z^N(x,t)\middle|\ S_0^N(t) = s^N_{0,i}, S_1^N(t) = s^N_{1,i}\right] &= \mathbb{E}\left[Z_0^N(x,t)\middle|\ S_0^N(t) = s^N_{0,i}, S_1^N(t) = s^N_{1,i}\right] \\ &= \frac{1}{|s^N_{0,i}|}\int_{\mathcal{S}_{Y_1}} f_{Y_0}\left(\frac{x - y_1 s^N_{1,i}}{s^N_{0,i}}\right) f_{Y_1}(y_1)\,\mathrm{d}y_1,\end{aligned} \tag{7.16}$$

where $\mathcal{S}_{Y_1}$ denotes the support of the random variable Y_1. If $|s^N_{1,i}| > |s^N_{0,i}|$, then

$$\begin{aligned}\mathbb{E}\left[Z^N(x,t)\middle|\ S_0^N(t) = s^N_{0,i}, S_1^N(t) = s^N_{1,i}\right] &= \mathbb{E}\left[Z_1^N(x,t)\middle|\ S_0^N(t) = s^N_{0,i}, S_1^N(t) = s^N_{1,i}\right] \\ &= \frac{1}{|s^N_{1,i}|}\int_{\mathcal{S}_{Y_0}} f_{Y_1}\left(\frac{x - y_0 s^N_{0,i}}{s^N_{1,i}}\right) f_{Y_0}(y_0)\,\mathrm{d}y_0,\end{aligned} \tag{7.17}$$

being $\mathcal{S}_{Y_0}$ the support of Y_0. These integrals (7.16), (7.17), can be approximated using numerical integration schemes. For instance, Gaussian quadrature rules [172, pp. 39–41] are appropriate since the density functions $f_{Y_1}(y_1)$ and $f_{Y_0}(y_0)$ act as integration weights. In the case of (7.16) (the case of (7.17) is analogous), one considers orthogonal polynomials $\{P_j(y_1)\}_{j=0}^{\infty}$ (the degree of P_j is j) with respect to the inner product of $\mathrm{L}^2(\mathcal{S}_{Y_1})$ with weight function $f_{Y_1}(y_1)$:

$$\int_{\mathcal{S}_{Y_1}} P_j(y_1)P_k(y_1)f_{Y_1}(y_1)\,\mathrm{d}y_1 = 0, \quad j \neq k.$$

For example, if Y_1 has a Normal or Uniform distribution, then $\{P_j(y_1)\}_{j=0}^{\infty}$ corresponds to the family of Hermite and Legendre polynomials, respectively. In other cases where Y_1 does not possess a standard probability distribution from the Askey scheme, $\{P_j(y_1)\}_{j=0}^{\infty}$ may be constructed through a Gram-Schmidt orthogonalization procedure from the canonical basis $\{1, y_1, y_1^2, y_1^3, \ldots\}$. The integral (7.16) is approximated as follows:

$$\frac{1}{|s_{0,i}^N|}\int_{\mathcal{S}_{Y_1}} f_{Y_0}\left(\frac{x - y_1 s_{1,i}^N}{s_{0,i}^N}\right) f_{Y_1}(y_1)\,\mathrm{d}y_1 \approx \frac{1}{|s_{0,i}^N|}\sum_{j=1}^{Q} w_{1,j}^{(Q)} f_{Y_0}\left(\frac{x - y_{1,j}^{(Q)} s_{1,i}^N}{s_{0,i}^N}\right). \tag{7.18}$$

Here, Q is the finite degree of the quadrature, the nodes $\{y_{1,j}\}_{j=1}^{Q}$ are the zeros of $P_Q(y_1)$, and $\{w_{1,j}^{(Q)}\}_{j=1}^{Q}$ are the weights so that the integration rule is exact for polynomials of degree less than or equal to $2Q-1$

The nodes $\{y_{0,j}\}_{j=1}^{Q}$, $\{y_{1,j}\}_{j=1}^{Q}$, and the weights $\{w_{0,j}^{(Q)}\}_{j=1}^{Q}$, $\{w_{1,j}^{(Q)}\}_{j=1}^{Q}$, only depend on f_{Y_0} and f_{Y_1}, therefore they are independent of the realization $(s_{0,i}^N, s_{1,i}^N)$, the number of simulations M and the truncation order N. Hence, the nodes and the weights should always be computed once for all at the beginning.

As more regularity we demand on f_{Y_0} and f_{Y_1}, faster convergence the quadrature rule will achieve with Q. The extreme case appears when f_{Y_0} and f_{Y_1} are analytic real functions: the quadrature rule converges exponentially with Q and therefore the bias error $\theta_N(x,t) = f_{X(t)}(x) - f_{X^N(t)}(x)$ is not seriously affected by the deterministic error arisen from the finite degree Q. Nonetheless, numerical problems in the quadrature approximations may arise when the integrands are strongly peaked. In this case, the degree Q must be taken larger, otherwise the bias error may increase dramatically.

The procedure described is implemented in the form of Algorithm 3. The density function $f_{X^N(t)}(x)$ is approximated by combining Monte Carlo simulation

on $S_0^N(t)$, $S_1^N(t)$, and quadrature rule for integration with respect to Y_0–Y_1. The output function is denoted as $f_X^{N,M,Q}(x,t)$.

Algorithm 3 Estimation of $f_{X^N(t)}(x)$ using Monte Carlo simulation and quadrature rule for integration, Section 7.3.

Inputs: t_0; N; f_{Y_0} (smooth); f_{Y_1} (smooth); probability distribution of $A_0,\ldots,A_N$, $B_0,\ldots,B_N$; number M of realizations in the Monte Carlo procedure; degree Q of the Gaussian quadrature rule; t; and discretization vector x of the density domain. The zeros $\{y_{0,j}^{(Q)}\}_{j=1}^Q$ and $\{y_{1,j}^{(Q)}\}_{j=1}^Q$ of the Q-th degree orthogonal polynomials with respect to f_{Y_0} and f_{Y_1}, respectively, have been previously determined. The weights $\{w_{0,j}^{(Q)}\}_{j=1}^Q$ and $\{w_{1,j}^{(Q)}\}_{j=1}^Q$ of the quadrature rules with weight functions f_{Y_0} and f_{Y_1}, respectively, have been previously calculated.

1: $\Sigma \leftarrow 0$ ▷ Initialize the samples sum
2: **for** $i = 1,\ldots,M$ **do** ▷ Monte Carlo loop
3: Set a realization of $S_0^N(t)$ and $S_1^N(t)$
4: **if** $|S_0^N(t)| \geq |S_1^N(t)|$ **then**
5: $\Sigma \leftarrow \Sigma + \sum_{j=1}^Q w_{1,j}^{(Q)} Z_0^N(x,t)|_{Y_1=y_{1,j}^{(Q)}}$ ▷ Quadrature rule with weight f_{Y_1}
6: **else**
7: $\Sigma \leftarrow \Sigma + \sum_{j=1}^Q w_{0,j}^{(Q)} Z_1^N(x,t)|_{Y_0=y_{0,j}^{(Q)}}$ ▷ Quadrature rule with weight f_{Y_0}
8: **end if**
9: **end for**
10: $f_X^{N,M,Q}(x,t) \leftarrow \Sigma/M$ ▷ Set sample average
11: **Return** $f_X^{N,M,Q}(x,t)$ ▷ Approximation of $f_{X^N(t)}(x)$

The complexity of Algorithm 3 is $\mathcal{O}(M(N^2+Q))$, in general. Recall that one realization of $S_0^N(t)$ and $S_1^N(t)$ requires $\mathcal{O}(N^2)$ operations when $A(t)$ and $B(t)$ are given by infinite random expansions. The quadrature rule demands $\mathcal{O}(Q)$ operations.

When $A(t)$ and $B(t)$ are random polynomials, one realization of $S_0^N(t)$ and $S_1^N(t)$ requires $\mathcal{O}(N)$ operations only. The complexity of the whole algorithm thus becomes $\mathcal{O}(M(N+Q))$.

In the case of simple problems (3.1) in which $S_0(t)$ and $S_1(t)$ are exactly known, Algorithm 3 is run with $N=\infty$, that is, by computing exact realizations of $S_0(t)$ and $S_1(t)$. The error in this case comes from the Monte Carlo simulation and the finite degree of the quadrature rule. The total complexity is given by $\mathcal{O}(MQ)$.

We finish this section with numerical experiments concerning the methodology and Algorithm 3. In the following three examples, we will compare Algorithm 2 and Algorithm 3 in terms of variance reduction, density approximation and error against complexity. The first two examples deal with problems having closed-form expressions for the fundamental sets ($N = \infty$), while the third example deals with orders of truncation $N < \infty$. After the three examples, we will conclude the section with recommendations regarding the use of Algorithm 2 or Algorithm 3.

Example 7.8 Consider (7.10) with input data $B \sim \text{Uniform}(1, 1.01)$ and $Y_0, Y_1 \sim \text{Normal}(1, \sigma = 2)$, all of them independent. As the fundamental set $\{S_0(t), S_1(t)\}$ is known with closed-form expressions, $S_0(t) = \cos(Bt)$ and $S_1(t) = \sin(Bt)/B$, we take $N = \infty$, by sampling from $S_0(t)$ and $S_1(t)$ directly.

Here the randomness mainly comes from the initial conditions. This is checked using the Sobol indices from (7.14) and (7.15). As proposed in [139, 159], the Sobol indices are computable using Monte Carlo simulation with the formulas

$$\mathbb{V}\left[Z_{S_0(t),S_1(t)}\right] \approx \frac{1}{M}\sum_{i=1}^{M} Z(s_i^I, y_i^I)Z(s_i^I, y_i^{II}) - \frac{1}{M}\sum_{i=1}^{M} Z(s_i^I, y_i^I)Z(s_i^{II}, y_i^{II}), \tag{7.19}$$

$$\mathbb{V}\left[Z_{Y_0,Y_1}\right] \approx \frac{1}{M}\sum_{i=1}^{M} Z(s_i^I, y_i^I)Z(s_i^{II}, y_i^I) - \frac{1}{M}\sum_{i=1}^{M} Z(s_i^I, y_i^I)Z(s_i^{II}, y_i^{II}), \tag{7.20}$$

where $\{s_i^I\}_{i=1}^M \subseteq \mathbb{R}^2$ and $\{s_i^{II}\}_{i=1}^M \subseteq \mathbb{R}^2$ denote two sets of joint realizations of $(S_0(t), S_1(t))$, and $\{y_i^I\}_{i=1}^M \subseteq \mathbb{R}^2$ and $\{y_i^{II}\}_{i=1}^M \subseteq \mathbb{R}^2$ denote two sets of joint realizations of (Y_0, Y_1), being all of them independent[1]. The notation used here, for instance $Z(s_i^I, y_i^I)$, means evaluating $Z(x,t)$ (given by (7.5)) at those specific realizations. In Figure 7.10, we report the estimates of the Sobol indices for $t = 10$. We observe that $\mathbb{S}_{Y_0,Y_1}$ is very close to 1, therefore indicating that the variability of $Z(x,t)$ is dominated by the initial conditions.

[1]If f is any function and U and V are independent random vectors, then $\mathbb{V}[\mathbb{E}[f(U,V)|V]] = \mathbb{E}[\mathbb{E}[f(U,V)|V]^2] - \mathbb{E}[\mathbb{E}[f(U,V)|V]]^2$ can be estimated using Monte Carlo simulation in the following expressions:

$$\mathbb{E}[\mathbb{E}[f(U,V)|V]^2] = \mathbb{E}_V[\mathbb{E}_U[f(U,V)]^2] = \mathbb{E}_V[\mathbb{E}_{U,U'}[f(U,V)f(U',V)]] = \mathbb{E}_{U,U',V}[f(U,V)f(U',V)],$$

$$\mathbb{E}[\mathbb{E}[f(U,V)|V]]^2 = \mathbb{E}_V[\mathbb{E}_U[f(U,V)]]^2 = \mathbb{E}_{U,V}[f(U,V)]^2 = \mathbb{E}_{U,V,U',V'}[f(U,V)f(U',V')],$$

where $U \overset{d}{=} U'$, $V \overset{d}{=} V'$, and U, V, U', V' are independent. The subindices for $\mathbb{E}$ indicate with respect to which random vectors the expectation is computed. This is a straightforward justification of (7.19) and (7.20). Notice that we do not estimate $\mathbb{E}[\mathbb{E}[f(U,V)|V]]^2$ with a sample average estimate for $\mathbb{E}[f(U,V)]^2$; according to [159, p. 959], [138, p. 284], the procedure we showed here yields better estimates for Sobol indices.

Thus we expect Algorithm 3 to be superior to Algorithm 2 in terms of variance reduction.

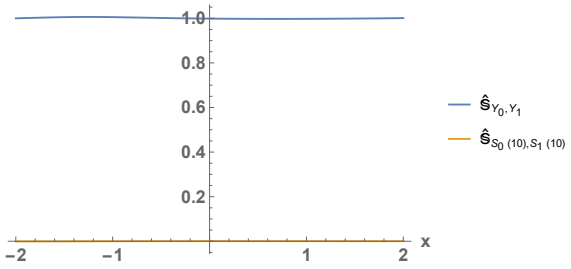

Figure 7.10: Estimated Sobol indices of the random inputs for $Z(x, t = 10)$. This figure corresponds to Example 7.8.

In this example, the Gaussian quadrature rule is based on Hermite polynomials. We work at time $t = 10$, with degree of quadrature $Q = 30$. Essentially, the quadrature rule approximation is error-free. In Figure 7.11 we plot estimates $\hat{\mathbb{V}}[Z(x, t = 10)]$ and $\hat{\mathbb{V}}[\mathbb{E}[Z(x, t = 10)|S_0(t = 10), S_1(t = 10)]]$. Observe that the magnitude of $\hat{\mathbb{V}}[\mathbb{E}[Z(x, t = 10)|S_0(t = 10), S_1(t = 10)]]$ is smaller than $\hat{\mathbb{V}}[Z(x, t = 10)]$, due to the law of total variance (7.13). Anyway, none of the variances becomes large at any point x, therefore good estimates of $f_{X(t=10)}(x)$ are obtained with any of the two algorithms.

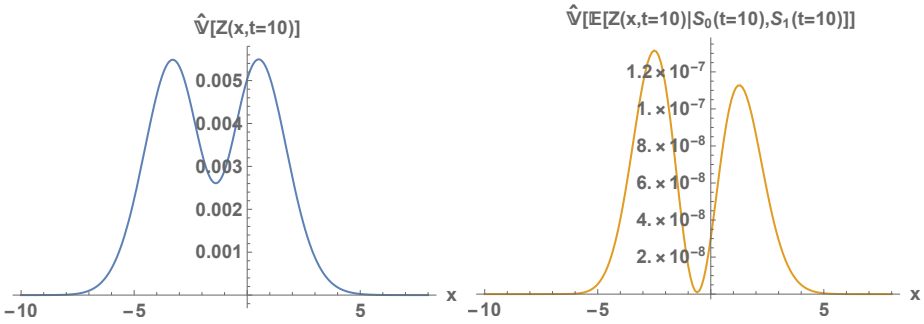

Figure 7.11: Estimates of $\hat{\mathbb{V}}[Z(x, t = 10)]$ and $\hat{\mathbb{V}}[\mathbb{E}[Z(x, t = 10)|S_0(t = 10), S_1(t = 10)]]$. This figure corresponds to Example 7.8.

In Figure 7.12 we depict the estimates for $f_{X(t=10)}(x)$ with Algorithm 2 (legend "Density MC") and Algorithm 3 (legend "Density MCQ"). Very simi-

lar smooth estimates are obtained, in fact they overlap, with $Q = 30$ and $M = 10,000$ realizations. We also compare the two estimates with the exact density function (legend "Exact density"). The exact density function is computable using a cubature rule for the double integral (6.1) (cubature rule constructed using a tensor product approach based on univariate Gauss-Hermite and Gauss-Legendre quadrature rules), because the integrand is smooth with no peaks (see Figure 7.13), thus posing no difficulties.

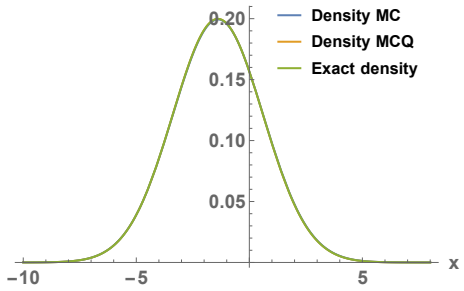

Figure 7.12: Estimates of $f_{X(t=10)}(x)$ with Monte Carlo simulation (MC), Monte Carlo simulation plus quadrature rule (MCQ), and exact density function. This figure corresponds to Example 7.8.

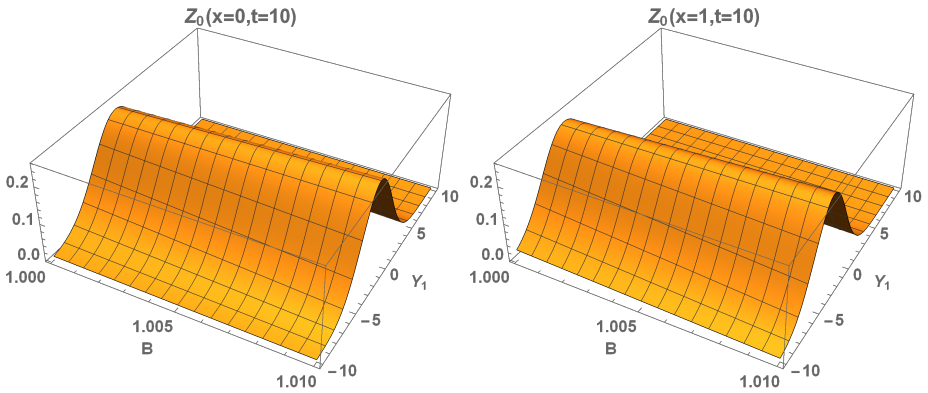

Figure 7.13: 3D plots of the integrand $Z_0(x, t = 10)$ at different points x. This figure corresponds to Example 7.8.

A further analysis is performed by studying the relation between the error and the complexity of Algorithm 2 and Algorithm 3. In Figure 7.14 we show

plots of mean relative error versus complexity in log-log scale for Algorithm 2 (legend "MC") and Algorithm 3 (legend "MCQ"), at different points x. Here we consider the complexity of Algorithm 2 as M, while the complexity of Algorithm 3 as MQ, being $Q = 30$ (recall that M denotes the number of realizations). The relative error is defined as the absolute error divided by the value of the exact density function. We run both algorithms up to complexity $\text{floor}(1.4^{44})$, and we keep the relative errors for complexities $\text{floor}(1.4^i)$, $i = 14, 15, \ldots, 44$ (a geometric progression makes equidistant complexities in log-scale). This gives a discretized realizable path of relative errors. We repeat this demanding process $n = 20$ times with parallel computing to obtain an estimate of the mean relative error (so that the analysis performed does not depend upon a specific realizable relative error). A 95% confidence interval of Gaussian-type for the mean relative error is constructed (this is justified by the central limit theorem), as mean plus/minus 1.96 times the standard error, where the standard error is the standard deviation of the original sample of length n divided by $\sqrt{n}$ [2]. The relative error descends with slope $1/2$ approximately, because the error of the Monte Carlo simulation goes down with M at rate $1/\sqrt{M}$. The mean relative error of Algorithm 3 is shifted down the mean relative error of Algorithm 2, due to the variance reduction of the estimators, thus indicating higher efficiency. Similar plots have been obtained for the points $x \in \{-1, 3\}$. Indeed, the effect of the initial conditions on the variability of Z is homogeneous in x, according to the Sobol indices from Figure 7.10, therefore a similar performance of the algorithm is expected for the different x's.

Example 7.9 We consider the same setting as Example 7.6: (7.10) with $B \sim \text{Uniform}(1, 2.5)$ and $Y_0, Y_1 \sim \text{Normal}(1, \sigma = 0.1)$, all of them independent.

In this case, the variability of Z is mainly due to B. This is deduced from Figure 7.15, where we report the Sobol indices (7.14) and (7.15) for $t = 10$, using the Monte Carlo estimates (7.19) and (7.20). The variance decomposition of Z is dominated by the effect of the fundamental set and the interactions, being the estimated Sobol index corresponding to the initial conditions very close to

[2] Here n might not be that large to apply the central limit theorem. Other types of confidence intervals may be considered, for instance those constructed via bootstrapping [61] with b samples with replacement from the original sample of length n. The percentile confidence interval has as left endpoint the empirical 0.025 percentile, and as right endpoint the empirical 0.975 percentile [61, p. 170], [55, p. 203]. The pivotal confidence interval generated by bootstrapping has left endpoint given by twice the mean minus the empirical 0.975 percentile, and right endpoint given by twice the mean minus the empirical 0.025 percentile [55, p. 194]. Finally, another Gaussian-type confidence interval may consider the mean plus/minus 1.96 times the bootstrap standard deviation of the mean. In all our examples, we checked that these confidence intervals are coinciding, thus showing robustness of our estimates. As n increases, better estimates for the mean are obtained.

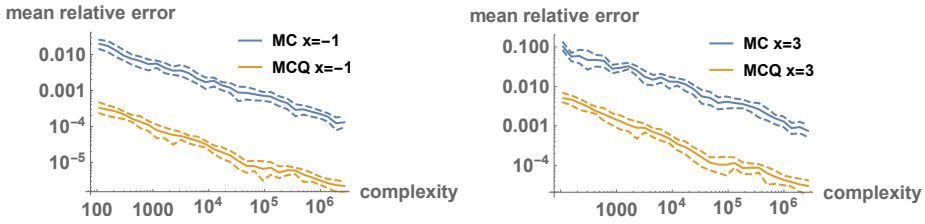

Figure 7.14: Plots in log-log scale of mean relative error versus complexity, at $t = 10$. The dashed lines represent 95% confidence intervals. This figure corresponds to Example 7.8.

0. Therefore, we do not expect much variance reduction from the quadrature rule.

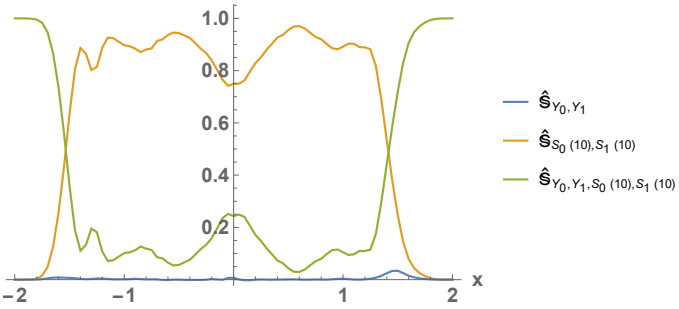

Figure 7.15: Estimated Sobol indices of the random inputs for $Z(x, t = 10)$. This figure corresponds to Example 7.9.

We work at $t = 10$, with degree of quadrature $Q = 30$ and Hermite polynomials. In Figure 7.16 we plot the variances of the statistical estimators from Algorithm 2 and Algorithm 3. We observe that the variances are greater than in Example 7.8. Also, in contrast with Example 7.8, the variance is not much reduced now with the quadrature rule. This was already expected when analyzing the effect of the inputs on the variability of Z using Sobol indices.

Figure 7.17 depicts the estimated density function $f_{X(t=10)}(x)$ with Algorithm 2 (legend "MC") and Algorithm 3 (legend "MCQ"), by setting $M = 40,000$ realizations and $Q = 30$. Both algorithms show similar estimates.

Figure 7.18 reports the log-log relationship between the relative error and the complexity at $t = 10$, for different points x. The setting is analogous to

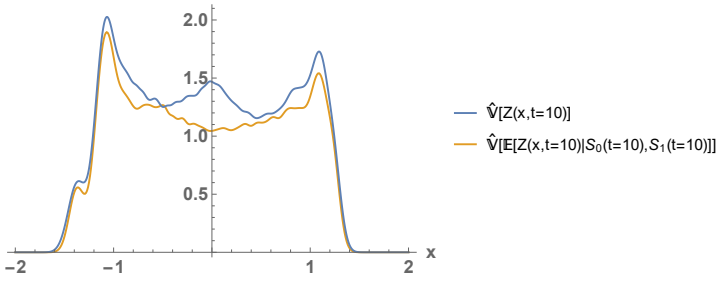

Figure 7.16: Estimates of $\hat{\mathbb{V}}[Z(x,t=10)]$ and $\hat{\mathbb{V}}[\mathbb{E}[Z(x,t=10)|S_0(t=10),S_1(t=10)]]$. This figure corresponds to Example 7.9.

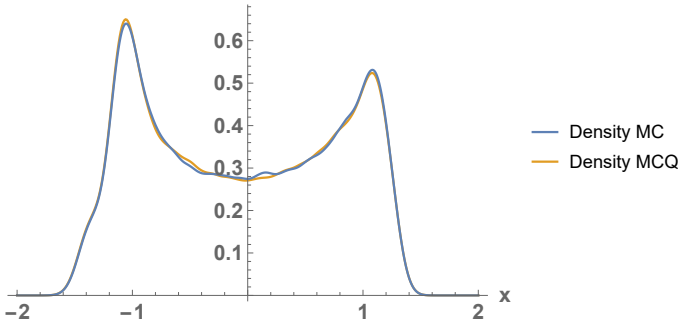

Figure 7.17: Estimates of $f_{X(t=10)}(x)$ with Monte Carlo simulation (MC), and Monte Carlo simulation plus quadrature rule (MCQ). This figure corresponds to Example 7.9.

Example 7.8. We consider the complexities of Algorithm 2 and Algorithm 3 as M and MQ, respectively, where $Q = 30$ and M is the number of realizations. The "absolute error" is calculated with respect to the density function obtained from Algorithm 2 with $100,000,000$ realizations (this is considered as our "exact density function"). We define the relative error as such absolute error divided by the value of the "exact density function". We run both algorithms up to complexity $\text{floor}(1.4^{44})$, while keeping the relative errors for complexities $\text{floor}(1.4^i)$, $i = 14, 15, \ldots, 44$. The procedure is repeated $n = 20$ times to determine an estimate for the mean relative error. A 95% Gaussian confidence interval for the mean relative error is constructed. See Example 7.8 for further details. We observe that the relative errors decay approximately with slope $1/2$, due to the rate of convergence proportional to $1/\sqrt{M}$ of the Monte Carlo simulation. In this example, the decay rate with slope $1/2$ is true for both algorithms as long as the sampling error of the "exact density function" does not become dominant. In the figure, we observe that the estimated relative error of Algorithm 2 tends to be constant sometime after 10^6 complexity, as it becomes of similar order of magnitude to the sampling error of the "exact density function". Algorithm 2 shows superior to Algorithm 3 in this example: in the plots, the depicted relation between the error and the complexity of Algorithm 3 is shifted to the right, because of the multiplication by Q of the number of realizations M in the complexity. As the effect of the initial conditions on the variability of Z is small, the potentiality of Algorithm 3 is lessened.

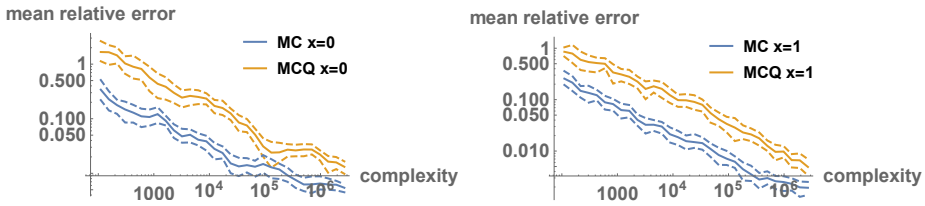

Figure 7.18: Plots in log-log scale of mean relative error versus complexity, at $t = 10$. The dashed lines represent 95% confidence intervals. This figure corresponds to Example 7.9.

In the two Examples 7.8, 7.9 presented until now, we studied two cases. When Y_0–Y_1 produce high variability on $\mathbb{V}[Z(x,t)]$, then Algorithm 3 clearly wins Algorithm 2 in terms of error versus complexity. On the contrary, when the dispersion of Y_0–Y_1 is small compared to $S_0(t)$ and $S_1(t)$, Algorithm 3 does no longer beat Algorithm 2. But in general, when the cost of obtaining realizations of $S_0(t)$ and $S_1(t)$ is larger (because we are working with truncated

power series in N, for instance), the cost due to the Monte Carlo procedure alone is comparable with the whole complexity. In such a case, Algorithm 2 never defeats Algorithm 3 by a big difference. This is the case studied in the following Example 7.10.

Example 7.10 Consider (7.11) with inputs $A_0 = 4$, $A_1 \sim \text{Uniform}(0,1)$, $B_0 \sim \text{Triangular}(-1,0)$, $B_1 \sim \text{Bernoulli}(0.35)$ and $Y_0, Y_1 \sim \text{Normal}(1, \sigma = 0.1)$, all of them independent. The random coefficients A_0, A_1, B_0 and B_1 are bounded and the initial conditions Y_0 and Y_1 have finite absolute moments, therefore there is an analytic solution $X(t)$ in the $\mathrm{L}^p(\Omega)$ sense, $1 \leq p < \infty$.

In this example, the fundamental set cannot be written with simple closed-form formulas. One needs to perform a dimensionality reduction of the problem by truncating the random power series at certain order N. We work at time $t = 1.5$ with order of truncation $N = 15$. The goal is to compare the approximations to the density $f_{X^{N=15}(t=1.5)}(x)$ offered by Algorithm 2 and Algorithm 3.

In Figure 7.19 we show the Sobol indices (7.14) and (7.15) for $t = 1.5$, computed with the Monte Carlo estimates (7.19) and (7.20). The Sobol indices demonstrate that the effect of Y_0–Y_1 on the variability of $Z^{N=15}(x, t = 1.5)$ is very small.

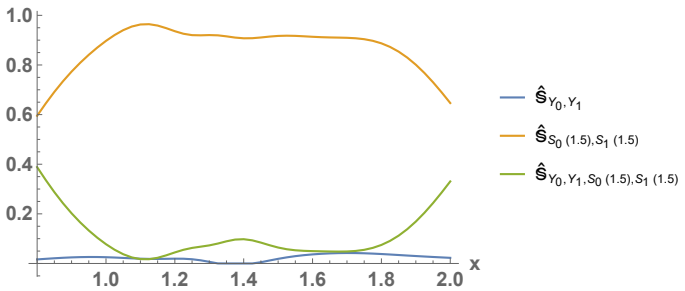

Figure 7.19: Estimated Sobol indices of the random inputs for $Z^{N=15}(x, t = 1.5)$. This figure corresponds to Example 7.10.

Hence, the variance reduction due to the quadrature rule is not very significant. This is confirmed by Figure 7.20, first panel, where we depict the variances of the statistical estimators. We use quadrature degree $Q = 30$, so that the expectation with respect to Y_0–Y_1 is practically error-free. In the sec-

ond plot, we estimate the density function $f_{X^{N=15}(t=1.5)}(x)$ using Algorithm 2 and Algorithm 3. We generated $M = 10,000$ realizations.

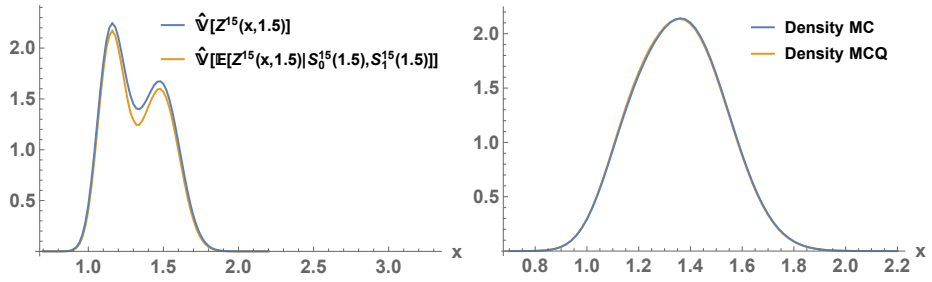

Figure 7.20: Estimates of $\hat{\mathbb{V}}[Z^{15}(x, t = 1.5)]$ and $\hat{\mathbb{V}}[\mathbb{E}[Z^{15}(x, t = 1.5)|S_0(t = 1.5), S_1(t = 1.5)]]$. Estimates of $f_{X^{15}(t=1.5)}(x)$ (the legend MC stands for the Monte Carlo method, the legend MCQ refers to Monte Carlo plus quadrature rule). This figure corresponds to Example 7.10.

An analogous analysis to Example 7.9 of relative error versus complexity is carried out here. The complexity of Algorithm 2 is MN, whereas the complexity of Algorithm 3 is $M(N+Q)$. Recall that M is the number of realizations, N is the truncation order for the power series, and Q is the quadrature degree. We work at the point $x = 1.4$, which belongs to the density domain (any other point is equally valid). We consider as our "exact density function" the output of Algorithm 2 with $M = 10,000,000$ realizations, i.e., complexity $MN = 150,000,000$. We run both algorithms for complexities floor(1.4^{37}), and keep the relative errors for complexities floor(1.4^i), $i = 14, 15, \ldots, 37$. The procedure is repeated $n = 20$ times (in parallel) to obtain the mean relative error. A 95% Gaussian confidence interval for the mean relative error has been constructed. Algorithm 3 is run with $Q = 10$ and $Q = 30$ degrees of quadrature. Figure 7.21 shows the results. Notice that, although Y_0–Y_1 have small dispersion compared with the fundamental set, both Algorithms show similar performances. Indeed, as each realization of the truncated fundamental set requires now more expense (cost N), Algorithm 2 cannot defeat Algorithm 3 by a big difference. The relative errors decrease with slope $1/2$, but for $Q = 10$ we perceive constant error for complexities greater than 10^5. This is because the bias error caused by the quadrature rule of degree $Q = 10$ has been attained, in contrast to the case $Q = 30$. This reflects the importance of selecting a sufficiently large quadrature degree, especially when f_{Y_0} and f_{Y_1} are peaked by small dispersion of Y_0 and Y_1.

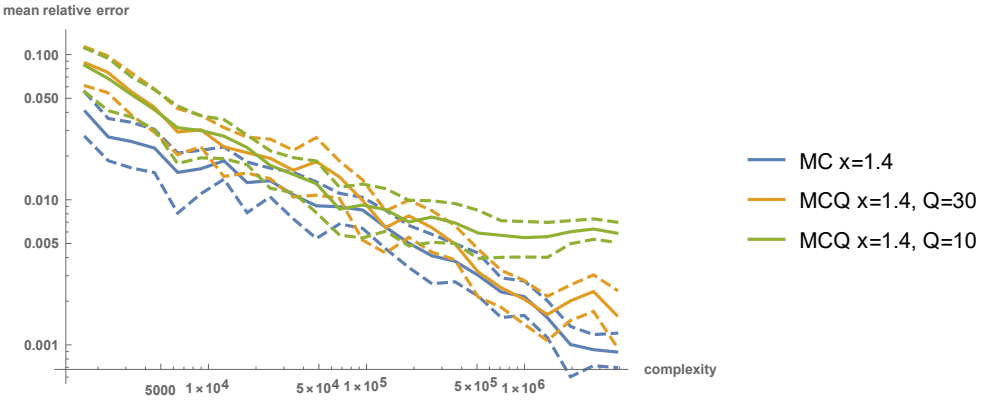

Figure 7.21: Plots in log-log scale of mean relative error versus complexity, at $t = 1.5$ and $N = 15$. The dashed lines represent 95% confidence intervals. This figure corresponds to Example 7.10.

The conclusion of this section is summarized as follows:

Case 1 When the dispersion of Y_0–Y_1 is large compared with $S_0(t)$ and $S_1(t)$ (this is checked with Sobol indices), then Algorithm 3 should be used. See Example 7.8.

Case 2 When the variability of Y_0–Y_1 is small compared with $S_0(t)$ and $S_1(t)$ (Sobol indices), and each realization of $S_0(t)$ and $S_1(t)$ takes certain computational burden (for instance, because we are dealing with truncated power series), then Algorithm 2 and Algorithm 3 are similar in terms of error versus complexity. See Example 7.10.

Case 3 Finally, if the variability of Y_0–Y_1 is small compared with $S_0(t)$ and $S_1(t)$ (Sobol indices), and each realization of $S_0(t)$ and $S_1(t)$ is immediately computed (when $S_0(t)$ and $S_1(t)$ are known with simple closed-form expressions), then Algorithm 2 beats Algorithm 3. See Example 7.9.

In any of the three cases, Q should be taken large enough to avoid a quadrature error greater than the sampling error.

7.4 Multilevel Monte Carlo simulation

The idea of multilevel Monte Carlo [75, 76] is to speed up the convergence of the Monte Carlo simulation by decomposing the target expectation into the sum of expectations of increments (through a telescopic sum identity), whose variances decay rapidly. In our case, we consider the following increments:

$$\Delta Z_0^l(x,t) = Z_0^{l+1}(x,t) - Z_0^l(x,t), \quad \Delta Z_1^l(x,t) = Z_1^{l+1}(x,t) - Z_1^l(x,t), \quad (7.21)$$

for each level $l \geq 0$, where each term was defined in (7.4). If we start from the level $l_0 \geq 0$, these increments yield the following telescopic sum identities:

$$\begin{aligned} f_{X(t)}(x) &= \mathbb{E}\left[Z_0^{l_0}(x,t)\right] + \sum_{l=l_0}^{\infty} \mathbb{E}\left[\Delta Z_0^l(x,t)\right] \\ &\approx f_{X^N(t)}(x) = \mathbb{E}\left[Z_0^{l_0}(x,t)\right] + \sum_{l=l_0}^{N-1} \mathbb{E}\left[\Delta Z_0^l(x,t)\right], \end{aligned}$$

$$\begin{aligned} f_{X(t)}(x) &= \mathbb{E}\left[Z_1^{l_0}(x,t)\right] + \sum_{l=l_0}^{\infty} \mathbb{E}\left[\Delta Z_1^l(x,t)\right] \\ &\approx f_{X^N(t)}(x) = \mathbb{E}\left[Z_1^{l_0}(x,t)\right] + \sum_{l=l_0}^{N-1} \mathbb{E}\left[\Delta Z_1^l(x,t)\right]. \end{aligned}$$

Each one of these expectations is approximated with Monte Carlo simulation. The rate of convergence of the Monte Carlo simulation depends on the number of realizations and $\mathbb{V}[Z_0^{l_0}(x,t)]$, $\mathbb{V}[\Delta Z_0^l(x,t)]$, and $\mathbb{V}[Z_1^{l_0}(x,t)]$, $\mathbb{V}[\Delta Z_1^l(x,t)]$. If $Z_0^l(x,t)$ or $Z_1^l(x,t)$ tend to $Z^l(x,t)$ in the mean square sense, then the variances of the increments, $\mathbb{V}[\Delta Z_0^l(x,t)]$, $\mathbb{V}[\Delta Z_1^l(x,t)]$, converge to 0, which fastens the convergence of the Monte Carlo procedure. The first level l_0 is usually chosen satisfying that, for $l \geq l_0$, the variances of the increments per level l are decaying monotonically and rapidly.

However, as we are selecting either f_{Y_0} or f_{Y_1} (as in Chapter 6), this may lead to numerical instabilities due to small denominators and large or infinite variance, which causes noisy features and slows down the convergence of the Monte Carlo procedure. This may hide the potentiality of multilevel Monte Carlo.

To sort out this issue, we path-wise select $\Delta Z_0^l(x,t)$ or $\Delta Z_1^l(x,t)$, by discarding the increment with the lowest denominator. The minimum among the denominators is considered as a time-dependent random variable

$$V^l(t) = \min\left\{|S_0^{l+1}(t)|, |S_0^l(t)|, |S_1^{l+1}(t)|, |S_1^l(t)|\right\}. \quad (7.22)$$

We pick the index with no smallest denominator through the random variable

$$U^l(t) = \{1 - i : \text{the minimum of } V^l(t) \text{ is attained at } i \in \{0, 1\}\}. \tag{7.23}$$

This random variable gives rise to two events,

$$H_0^l(t) = \{U^l(t) = 0\}, \quad H_1^l(t) = \{U^l(t) = 1\}.$$

Let

$$Z_\Delta^l(x,t) = \Delta Z_0^l(x,t)\mathbb{1}_{H_0^l(t)} + \Delta Z_1^l(x,t)\mathbb{1}_{H_1^l(t)}.$$

This random process models the selection of $\Delta Z_0^l(x,t)$ and $\Delta Z_1^l(x,t)$, according to whether their expressions do not have the smallest denominator.

Let us see that with the expectations of $Z_\Delta^l(x,t)$ we recover $f_{X(t)}(x)$, by summing up to infinity, and $f_{X^N(t)}(x)$, by summing up to $N-1$.

Proposition 7.11 *We have the following sum identities for the density functions $f_{X(t)}(x)$ and $f_{X^N(t)}(x)$:*

$$f_{X(t)}(x) = \mathbb{E}\left[Z^{l_0}(x,t)\right] + \sum_{l=l_0}^{\infty} \mathbb{E}\left[Z_\Delta^l(x,t)\right],$$

$$f_{X^N(t)}(x) = \mathbb{E}\left[Z^{l_0}(x,t)\right] + \sum_{l=l_0}^{N-1} \mathbb{E}\left[Z_\Delta^l(x,t)\right].$$

Proof. First, by Proposition 7.2, $\mathbb{E}[Z^{l_0}(x,t)] = f_{X^{l_0}(t)}(x)$. On the other hand,

$$\begin{aligned}
\mathbb{E}[Z_\Delta^l(x,t)] &= \mathbb{E}\left[\Delta Z_0^l(x,t)\mathbb{1}_{H_0^l(t)}\right] + \mathbb{E}\left[\Delta Z_1^l(x,t)\mathbb{1}_{H_1^l(t)}\right] \\
&= \mathbb{E}\left[\Delta Z_0^l(x,t)\,\middle|\,H_0^l(t)\right]\mathbb{P}\left[H_0^l(t)\right] + \mathbb{E}\left[\Delta Z_1^l(x,t)\,\middle|\,H_1^l(t)\right]\mathbb{P}\left[H_1^l(t)\right] \\
&= \mathbb{E}\left[Z_0^{l+1}(x,t)\,\middle|\,H_0^l(t)\right]\mathbb{P}\left[H_0^l(t)\right] - \mathbb{E}\left[Z_0^l(x,t)\,\middle|\,H_0^l(t)\right]\mathbb{P}\left[H_0^l(t)\right] \\
&\quad + \mathbb{E}\left[Z_1^{l+1}(x,t)\,\middle|\,H_1^l(t)\right]\mathbb{P}\left[H_1^l(t)\right] - \mathbb{E}\left[Z_1^l(x,t)\,\middle|\,H_1^l(t)\right]\mathbb{P}\left[H_1^l(t)\right].
\end{aligned}$$

By Lemma 7.1 and the law of total probability,

$$\begin{aligned}
\mathbb{E}[Z_\Delta^l(x,t)] &= f_{X^{l+1}(t)}\left(x\,\middle|\,H_0^l(t)\right)\mathbb{P}\left[H_0^l(t)\right] - f_{X^l(t)}\left(x\,\middle|\,H_0^l(t)\right)\mathbb{P}\left[H_0^l(t)\right] \\
&\quad + f_{X^{l+1}(t)}\left(x\,\middle|\,H_1^l(t)\right)\mathbb{P}\left[H_1^l(t)\right] - f_{X^l(t)}\left(x\,\middle|\,H_1^l(t)\right)\mathbb{P}\left[H_1^l(t)\right] \\
&= f_{X^{l+1}(t)}(x) - f_{X^l(t)}(x).
\end{aligned}$$

The proposition follows from the telescopic sum identities $f_{X(t)}(x) = f_{X^{l_0}(t)}(x) + \sum_{l=l_0}^{\infty}(f_{X^{l+1}(t)}(x) - f_{X^l(t)}(x))$ and $f_{X^N(t)}(x) = f_{X^{l_0}(t)}(x) + \sum_{l=l_0}^{N-1}(f_{X^{l+1}(t)}(x) - f_{X^l(t)}(x))$.

□

With the objective of approximating the target density function $f_{X(t)}(x)$ pointwise, we truncate the telescopic sum identity at increment $N-1$, which causes the bias error $\theta_N(x,t) = f_{X(t)}(x) - f_{X^N(t)}(x)$:

$$f_{X(t)}(x) \approx f_{X^N(t)}(x) = \mathbb{E}\left[Z^{l_0}(x,t)\right] + \sum_{l=l_0}^{N-1} \mathbb{E}\left[Z_\Delta^l(x,t)\right].$$

The expectations $\mathbb{E}[Z^{l_0}(x,t)]$, $\mathbb{E}[Z_\Delta^l(x,t)]$, $l = l_0, \ldots, N-1$, are computed independently with Monte Carlo simulation, whose rate of convergence depends on the finite number of realizations and on $\mathbb{V}[Z^{l_0}(x,t)]$, $\mathbb{V}[Z_\Delta^l(x,t)]$, $l = l_0, \ldots, N-1$.

We have

$$\left\|Z_\Delta^l(x,t)\right\|_2 \leq \left\|\Delta Z_0^l(x,t)\right\|_2 + \left\|\Delta Z_1^l(x,t)\right\|_2.$$

Thus, if both $\Delta Z_0^l(x,t)$ and $\Delta Z_1^l(x,t)$ have finite variances, then so does $Z_\Delta^l(x,t)$. On the contrary, if $\Delta Z_0^l(x,t)$ or $\Delta Z_1^l(x,t)$ have infinite variances, then $Z_\Delta^l(x,t)$ may have finite variance, instead. In any of these cases, we cannot get worse in terms of infinite variance.

We hope to have the variances $\mathbb{V}[Z^{l_0}(x,t)]$ and $\mathbb{V}[Z_\Delta^l(x,t)]$, $l = l_0, \ldots, N-1$, finite, to exploit the decay of $\mathbb{V}[Z_\Delta^l(x,t)]$ to 0 with l. The first level l_0 is chosen such that $\mathbb{V}[Z_\Delta^l(x,t)]$ is decreasing with $l \geq l_0$, monotonically and fast. The best case is when $\log \mathbb{V}[Z_\Delta^l(x,t)]$ is decreasing linearly with $l \geq l_0$, as exponential decay holds for the variances of the increments.

Let $M = (N_{l_0}; M_{l_0}, \ldots, M_{N-1})$ be the number of simulations performed: N_{l_0} for $Z^{l_0}(x,t)$, M_l for $Z_\Delta^l(x,t)$, $l = l_0, \ldots, N-1$. In principle, the number of simulations may depend on (x,t), although we drop this possible dependence along this discussion.

Let $f_X^{N,M}(x,t)$ be the final approximation to $f_{X(t)}(x)$:

$$f_X^{N,M}(x,t) = \frac{1}{N_{l_0}} \sum_{i=1}^{N_{l_0}} Z^{l_0}(x,t,\omega_{i,l_0}) + \sum_{l=l_0}^{N-1} \frac{1}{M_l} \sum_{i=1}^{M_l} Z_\Delta^l(x,t,\omega_{i,l,\Delta}), \qquad (7.24)$$

where ω with the appropriate subindices denote outcomes. The mean square error in the approximation to $f_{X^N(t)}(x)$ is given by the variance of the statis-

tical error $\mathcal{E}_{N,M}(x,t) = f_{X^N(t)}(x) - f_X^{N,M}(x,t)$, which is

$$\left\| f_{X^N(t)}(x) - f_X^{N,M}(x,t) \right\|_2^2 = \mathbb{V}\left[\mathcal{E}_{N,M}(x,t)\right] = \frac{\mathbb{V}\left[Z^{l_0}(x,t)\right]}{N_{l_0}} + \sum_{l=l_0}^{N-1} \frac{\mathbb{V}\left[Z_\Delta^l(x,t)\right]}{M_l}. \tag{7.25}$$

The total cost is

$$W_{N,M} = N_{l_0} c_{l_0} + \sum_{l=l_0}^{N-1} M_l C_l, \tag{7.26}$$

where c_{l_0} denotes the number of operations per realization of $Z^{l_0}(x,t)$, and C_l denotes the number of operations per realization of $Z_\Delta^l(x,t)$, $l = l_0, \ldots, N-1$. In our case, we consider $c_l = C_l = l^2$ when dealing with $A(t)$ and $B(t)$ as random infinite expansions, and $c_l = C_l = l$ when $A(t)$ and $B(t)$ are random polynomials.

Proposition 7.12 *Given a fixed variance* (7.25) ϵ^2*, the cost* (7.26) *is minimized by selecting*

$$N_{l_0}(x,t) = \epsilon^{-2} \left(\sqrt{\mathbb{V}\left[Z^{l_0}(x,t)\right] c_{l_0}} + \sum_{l=l_0}^{N-1} \sqrt{\mathbb{V}\left[Z_\Delta^l(x,t)\right] C_l} \right) \sqrt{\frac{\mathbb{V}\left[Z^{l_0}(x,t)\right]}{c_{l_0}}}, \tag{7.27}$$

$$M_l(x,t) = \epsilon^{-2} \left(\sqrt{\mathbb{V}\left[Z^{l_0}(x,t)\right] c_{l_0}} + \sum_{l=l_0}^{N-1} \sqrt{\mathbb{V}\left[Z_\Delta^l(x,t)\right] C_l} \right) \sqrt{\frac{\mathbb{V}\left[Z_\Delta^l(x,t)\right]}{C_l}}, \tag{7.28}$$

for $l = l_0, \ldots, N-1$*. In such a case, the total cost* (7.26) *is given by*

$$W_{N,M} = \epsilon^{-2} \left(\sqrt{\mathbb{V}\left[Z^{l_0}(x,t)\right] c_{l_0}} + \sum_{l=l_0}^{N-1} \sqrt{\mathbb{V}\left[Z_\Delta^l(x,t)\right] C_l} \right)^2. \tag{7.29}$$

Proof. To simplify the notation, we drop the dependencies on t and x. By treating the integers N_{l_0} and M_l as continuous variables, we want to minimize the real function

$$W\left(N_{l_0}; M_{l_0}, \ldots, M_{N-1}\right) = N_{l_0} c_{l_0} + \sum_{l=l_0}^{N-1} M_l C_l$$

subject to the constrain

$$\mathbb{V}\left(N_{l_0}; M_{l_0}, \ldots, M_{N-1}\right) = \frac{\mathbb{V}\left[Z^{l_0}\right]}{N_{l_0}} + \sum_{l=l_0}^{N-1} \frac{\mathbb{V}\left[Z_\Delta^l\right]}{M_l} = \epsilon^2.$$

Using Lagrange multipliers, we consider the Lagrangian function $\phi = W + \lambda \mathbb{V}$. We impose $\nabla \phi = 0$ to find its stationary points:

$$0 = \frac{\partial \phi}{\partial N_{l_0}} = c_{l_0} - \lambda \frac{\mathbb{V}\left[Z^{l_0}\right]}{N_{l_0}^2}, \quad 0 = \frac{\partial \phi}{\partial M_l} = C_l - \lambda \frac{\mathbb{V}\left[Z_\Delta^l\right]}{M_l^2}.$$

We obtain the unique critical point

$$\tilde{N}_{l_0} = \mu \sqrt{\frac{\mathbb{V}\left[Z^{l_0}\right]}{c_{l_0}}}, \quad \tilde{M}_l = \mu \sqrt{\frac{\mathbb{V}\left[Z_\Delta^l\right]}{C_l}},$$

being $\mu = \sqrt{\lambda}$. Using the restriction on $\mathbb{V}$ we compute the parameter μ:

$$\mu = \epsilon^{-2} \left(\sqrt{\mathbb{V}\left[Z^{l_0}(x,t)\right] c_{l_0}} + \sum_{l=l_0}^{N-1} \sqrt{\mathbb{V}\left[Z_\Delta^l(x,t)\right] C_l} \right). \tag{7.30}$$

Hence, we obtain the stated number of simulations in (7.27)–(7.28).
Let us see that $\tilde{M} = (\tilde{N}_{l_0}; \tilde{M}_{l_0}, \ldots, \tilde{M}_{N-1})$ is indeed a global minimum of $\mathbb{V}$ on

$$\Theta = \left\{ M = (N_{l_0}; M_{l_0}, \ldots, M_{N-1}) \in (0,\infty)^{N-l_0+1} : \mathbb{V}(M) = \epsilon^2 \right\}.$$

First, notice that

$$N_{l_0} > \mathbb{V}\left[Z^{l_0}\right] \epsilon^{-2}, \quad M_l > \mathbb{V}\left[Z_\Delta^l\right] \epsilon^{-2}, \quad l = l_0, \ldots, N-1.$$

Second, since $\mathbb{V}(M) \to +\infty$ whenever any component of M tends to $+\infty$, we can find an interval $(0,R)^{N-l_0+1}$ such that $\mathbb{V}(M) \geq 2\mathbb{V}(\tilde{M})$ for any $M \notin (0,R)^N$, $M \in \Theta$. Hence, we can work in the compact set

$$\tilde{\Theta} = \Theta \cap \left(\left[\mathbb{V}\left[Z^{l_0}\right] \epsilon^{-2}, R\right] \times \prod_{l=l_0}^{N-1} \left[\mathbb{V}\left[Z_\Delta^l\right] \epsilon^{-2}, R\right] \right),$$

where $\mathbb{V}$ attains a global minimum by Weierstrass extreme value theorem.
Let $\overline{M}$ be such a global minimum. By the previous discussion,

$$\overline{M} \in \left(\mathbb{V}\left[Z^{l_0}\right] \epsilon^{-2}, R\right) \times \prod_{l=l_0}^{N-1} \left(\mathbb{V}\left[Z_\Delta^l\right] \epsilon^{-2}, R\right).$$

As $\overline{M}$ belongs to an open set, necessarily $\nabla \phi(\overline{M}, \overline{\lambda}) = 0$, for some parameter $\overline{\lambda}$. Then $\overline{M} = \tilde{M}$, as wanted.

□

For a truncation order N, and for a statistical error variance ϵ^2 by running $\epsilon^{-2}\mathbb{V}[Z^N(x,t)]$ simulations, the cost of Algorithm 2 is $\epsilon^{-2}\mathbb{V}[Z^N(x,t)]c_N$. For fixed N, both approaches (standard Monte Carlo in Algorithm 2 and multilevel Monte Carlo) yield a cost $\mathcal{O}(\epsilon^{-2})$. However, there is a significant difference when N grows. The complexity $c_N = N$ (if $A(t)$ and $B(t)$ are random polynomials) or $c_N = N^2$ (if $A(t)$ and $B(t)$ are infinite random series) per realization of $Z^N(x,t)$ grows to infinity. Then the cost of Algorithm 2 tends to infinity as $N \to \infty$, linearly (if $c_N = N$) or quadratically (if $c_N = N^2$). If we choose $N = \mathcal{O}(\log \epsilon^{-1})$ to ensure a bias error less than ϵ (under exponential convergence of $f_{X^N(t)}(x)$ with N), then the complexity of Algorithm 2 is $\mathcal{O}(\epsilon^{-2}\log^2\epsilon)$, when $c_N = N^2$, and $\mathcal{O}(\epsilon^{-2}\log\epsilon^{-1})$, when $c_N = N$. However, in the multilevel approach the cost (7.29) remains $\mathcal{O}(\epsilon^{-2})$ despite having $N \to \infty$, as $\sqrt{\mathbb{V}[Z^{l_0}(x,t)]c_{l_0}} + \sum_{l=l_0}^{\infty}\sqrt{\mathbb{V}[Z^l_\Delta(x,t)]C_l} < \infty$ in general. For instance, when $\mathbb{V}[Z^l_\Delta(x,t)]$ decreases exponentially with l. Hence, from this theoretical analysis, multilevel Monte Carlo is superior to standard Monte Carlo, at least when N grows.

In practice, multilevel Monte Carlo shows up less complexity than standard Monte Carlo from certain level l_0 and $N > l_0$. It is very convenient when it comes to optimality of the multilevel Monte Carlo approach to pick l_0 such that its complexity is strictly less than that of Algorithm 2. This choice can be made heuristically, on the fly, and it does not pose any challenge.

In numerical applications, the main computational burden of multilevel Monte Carlo is due to the number of samples N_{l_0} of $Z^{l_0}(x,t)$. For the last levels near N, which are the most complex ones, we run only few samples. In general, from a certain level l, only 1 realization of $Z^l_\Delta(x,t)$ is required. From a certain order of truncation N, the complexity of multilevel Monte Carlo will be constant; whereas the complexity of the standard Monte Carlo strategy will be growing up to infinity with N, linearly or quadratically. In our setting, multilevel Monte Carlo turns out to be useful when high accuracy is required.

Given l_0, $N > l_0$ and ϵ, the cost of multilevel Monte Carlo is approximately $\epsilon^{-2}\mathbb{V}[Z^{l_0}(x,t)]c_{l_0}$, because the sequence $\mathbb{V}[Z^l_\Delta(x,t)]C_l$ decreases rapidly with l. The expense of standard Monte Carlo is $\epsilon^{-2}\mathbb{V}[Z^N(x,t)]c_N$. Assuming that $\mathbb{V}[Z^{l_0}(x,t)] \approx \mathbb{V}[Z^N(x,t)]$, which is quite realistic if l_0 and $N > l_0$ are not too small, the cost of multilevel Monte Carlo is reduced by factor c_{l_0}/c_N, approximately. The largest reduction is thus obtained for problems (3.1) with infinite series inputs, since in such a case c_N grows quadratically with N.

The complete multilevel Monte Carlo approach is presented in what follows in the form of algorithms. Fixed the time t and the maximum level N, the whole procedure is divided into different parts:

Part 1 In Algorithm 4, we determine $\mathbb{V}[Z_\Delta^l(x,t)]$ for $0 \leq l \leq N-1$. They are computed *à la* Monte Carlo, but not using an excessive amount $m_\mathbb{V}$ of realizations, as we are only interested in their magnitudes. The main loop over the levels l can be carried out in parallel.

A possible approach to pick $m_\mathbb{V}$ is based on the following fact: the statistical error variance in the Monte Carlo estimation of $\mathbb{V}[Z_\Delta^l(x,t)]$ is given by

$$\frac{\mathbb{V}\left[\left(Z_\Delta^l - \mathbb{E}\left[Z_\Delta^l\right]\right)^2\right]}{m_\mathbb{V}} = \frac{\left(\kappa\left[Z_\Delta^l\right] - 1\right)\mathbb{V}[Z_\Delta^l(x,t)]^2}{m_\mathbb{V}},$$

where κ denotes the kurtosis. Thus, the accuracy of the Monte Carlo estimation of $\mathbb{V}[Z_\Delta^l(x,t)]$ depends on the kurtosis of $Z_\Delta^l(x,t)$. In practice, if desired, a preliminary analysis of the magnitude of $\kappa[Z_\Delta^l(x,t)]$ using Monte Carlo simulation could be carried out to select $m_\mathbb{V}$ properly.

Part 2 In Algorithm 5, we estimate $\mathbb{V}[Z^l(x,t)]$ for $1 \leq l \leq N$ (with parallel computing). Then we set the ratios between the complexities of Algorithm 2,

$$W_N^{\mathrm{MC}}(t) = \epsilon^{-2} c_N \max_x \mathbb{V}[Z^N(x,t)], \tag{7.31}$$

and of the multilevel Monte Carlo strategy with initial level $l_1 = 1, \ldots, N-1$,

$$W_{l_1,N}^{\mathrm{multi}}(t) = \epsilon^{-2} \max_x \left(\sqrt{\mathbb{V}\left[Z^{l_1}(x,t)\right] c_{l_1}} + \sum_{l=l_1}^{N-1} \sqrt{\mathbb{V}\left[Z_\Delta^l(x,t)\right] C_l}\right)^2. \tag{7.32}$$

These estimates are done via standard Monte Carlo simulation with not too many realizations $m_\mathbb{V}$. The ratio of the complexities is given by

$$R_{l_1,N}(t) = \frac{W_N^{\mathrm{MC}}(t)}{W_{l_1,N}^{\mathrm{multi}}(t)} = \frac{\max_x \mathbb{V}[Z^N(x,t)] c_N}{\max_x \left(\sqrt{\mathbb{V}\left[Z^{l_1}(x,t)\right] c_{l_1}} + \sum_{l=l_1}^{N-1} \sqrt{\mathbb{V}\left[Z_\Delta^l(x,t)\right] C_l}\right)^2}. \tag{7.33}$$

Notice that this ratio is independent of any statistical error variance ϵ^2. In order not to have x-dependent complexities, we take the maximum on x lying in the density domain of interest. If $R_{l_1,N}(t) > 1$, then multilevel Monte Carlo starting from level l_1 is more efficient than standard Monte Carlo.

We define the first level $l_0 = l_0(N,t)$ as the maximizer of

$$\max_{l_1:\, 1\le l_1<N} R_{l_1,N}(t) > 1. \tag{7.34}$$

This way we are achieving the maximum optimality of multilevel Monte Carlo. In general, the variances $\mathbb{V}[Z^l_\Delta(x,t)]$ start to decrease monotonically and fast with l before the maximizer l_0.

Part 3 Finally, in Algorithm 6, we perform the whole multilevel Monte Carlo procedure. Given a statistical error variance ϵ^2 (see (7.25)), and given the first level l_0 obtained in Part 2, we first determine the parameter μ using (7.30). Then we obtain the optimal $N_{l_0}(x,t)$ and $M_l(x,t)$ for $l = l_0,\dots,N-1$, using (7.27) and (7.28) from the constrained minimization problem. We set $N_{l_0}(t)$ and $M_l(t)$ as the ceiling of the maximum on x lying in the density domain of interest. Thus, the number of simulations is an integer independent of x.

Once the number of simulations is fixed, we estimate $\mathbb{E}[Z^{l_0}(x,t)]$ and $\mathbb{E}[Z^l_\Delta(x,t)]$, $l = l_0,\dots,N-1$, using standard Monte Carlo simulation. The output discretized function $f_X^{N,M}(x,t)$ is (7.24), which estimates $f_{X^N(t)}(x)$ with root mean square error $\|f_{X^N(t)}(x) - f_X^{N,M}(x,t)\|_2 \le \epsilon$, for all x.

In general, the statistical error variance ϵ^2 and the order of truncation N are chosen as follows. At the beginning, one can estimate the bias error of the density approximation: given a very large order of truncation $N_\infty \gg 1$, and taking into account that $Z^{N_\infty}(x,t) \approx Z(x,t)$, the bias error at order $1 \le N \ll N_\infty$, $|f_{X(t)}(x) - f_{X^N(t)}(x)|$, is estimated as

$$\left|\widehat{\mathbb{E}}\left[Z^{N_\infty}(x,t) - Z^N(x,t)\right]\right|. \tag{7.35}$$

The expectation is estimated using standard Monte Carlo simulation. Suppose we fix a target global error δ. We split the target error as the sum of the bias error and the sampling error: $\delta = \epsilon_{\text{bias}} + \epsilon_{\text{sampling}}$. We take N such that (7.35) is less than ϵ_{bias} for all x. Then we apply the multilevel Monte Carlo procedure with statistical error variance $\epsilon^2_{\text{sampling}}$. Keep in mind that the error δ cannot be as small as we want, since the sampling error $\epsilon_{\text{sampling}}$ must be achievable.

To apply the multilevel Monte Carlo procedure, we need N such that the maximum (7.34) is greater than 1. This holds for large N for sure, according to the theoretical discussion. Thus, in practice, we can choose ϵ_{bias} small to make N large, and then run the multilevel Monte Carlo algorithm with $\epsilon_{\text{sampling}} = \delta - \epsilon_{\text{bias}}$. In general, the decomposition of δ into ϵ_{bias} and $\epsilon_{\text{sampling}}$

Algorithm 4 Computation of $\mathbb{V}[Z_\Delta^l(x,t)]$ for $0 \leq l \leq N-1$, see Part 1 in Section 7.4.

Inputs: t_0; N; f_{Y_0}; f_{Y_1}; probability distribution of $A_0, \ldots, A_N$, $B_0, \ldots, B_N$; number $m_{\mathbb{V}}$ of realizations in the Monte Carlo procedure; t; and discretization vector x of the density domain.

for $l = 0, \ldots, N-1$ **do** ▷ Loop through levels
  $\Sigma_1, \Sigma_2 \leftarrow 0$ ▷ Initialize the samples sum
  for $i = 1, \ldots, m_{\mathbb{V}}$ **do** ▷ Monte Carlo loop
    Set a realization of $S_0^l(t)$, $S_1^l(t)$, $S_0^{l+1}(t)$ and $S_1^{l+1}(t)$
    if $U_l = 0$ **then** ▷ See (7.22)–(7.23)
      Set a realization of Y_1
      $\Sigma_1 \leftarrow \Sigma_1 + \Delta Z_0^l(x,t)$ ▷ See (7.21)
      $\Sigma_2 \leftarrow \Sigma_2 + \Delta Z_0^l(x,t)^2$
    else
      Set a realization of Y_0
      $\Sigma_1 \leftarrow \Sigma_1 + \Delta Z_1^l(x,t)$ ▷ See (7.21)
      $\Sigma_2 \leftarrow \Sigma_2 + \Delta Z_1^l(x,t)^2$
    end if
  end for
  $\mathbb{V}[Z_\Delta^l(x,t)] \leftarrow \Sigma_2/m_{\mathbb{V}} - (\Sigma_1/m_{\mathbb{V}})^2$ ▷ Set sample variance
end for
Return $\mathbb{V}[Z_\Delta^l(x,t)]$, $l = 0, \ldots, N-1$ ▷ Estimate of $\mathbb{V}[Z_\Delta^l(x,t)]$

Algorithm 5 Estimates of l_0 and $\mathbb{V}[Z^{l_0}(x,t)]$, see Part 2 in Section 7.4.

Inputs: t_0; N; f_{Y_0}; f_{Y_1}; probability distribution of $A_0,\dots,A_N$, $B_0,\dots,B_N$; variances $\mathbb{V}[Z^l_\Delta(x,t)]$, $l=0,\dots,N-1$, determined in Algorithm 4; number $m_{\mathbb{V}}$ of realizations in the Monte Carlo procedure; costs c_l and C_l, $l=0,\dots,N$; t; and discretization vector x of the density domain.

for $l=1,\dots,N$ **do** ▷ Loop through levels
  $\Sigma_1,\Sigma_2 \leftarrow 0$ ▷ Initialize the samples sum
  for $i=1,\dots,m_{\mathbb{V}}$ **do** ▷ Monte Carlo loop
    Set a realization of $S_0^l(t)$ and $S_1^l(t)$
    if $|S_0^l(t)| \geq |S_1^l(t)|$ **then**
      Set a realization of Y_1
      $\Sigma_1 \leftarrow \Sigma_1 + Z_0^l(x,t)$ ▷ See (7.4)
      $\Sigma_2 \leftarrow \Sigma_2 + Z_0^l(x,t)^2$
    else
      Set a realization of Y_0
      $\Sigma_1 \leftarrow \Sigma_1 + Z_1^l(x,t)$ ▷ See (7.4)
      $\Sigma_2 \leftarrow \Sigma_2 + Z_1^l(x,t)^2$
    end if
  end for
  $\mathbb{V}[Z^l(x,t)] \leftarrow \Sigma_2/m_{\mathbb{V}} - (\Sigma_1/m_{\mathbb{V}})^2$ ▷ Set sample variance
end for
Set ratio (7.33) of complexities, $R_{l_1,N}(t)$, $l_1=1,\dots,N-1$
Take l_0 as the maximizer of the ratio, between levels $1,\dots,N-1$ (see (7.34))
Return l_0 and $\mathbb{V}[Z^{l_0}(x,t)]$

Algorithm 6 Computation of $f_{X^{N(t)}}(x)$ with multilevel Monte Carlo simulation, see Part 3 in Section 7.4.

Inputs: t_0; N; f_{Y_0}; f_{Y_1}; probability distribution of $A_0, \ldots, A_N$, $B_0, \ldots, B_N$; first level l_0 determined on the fly from Algorithm 5; variances $\mathbb{V}[Z_\Delta^l(x,t)]$, $l = l_0, \ldots, N-1$, determined from Algorithm 4; variance $\mathbb{V}[Z^{l_0}(x,t)]$ determined from Algorithm 5; target root mean square error ϵ; costs c_{l_0} and C_l, $l = l_0, \ldots, N-1$; t; and discretization vector x of the density domain.

$\mu(x,t) \leftarrow \epsilon^{-2}(\sqrt{\mathbb{V}\left[Z^{l_0}(x,t)\right] c_{l_0}} + \sum_{l=l_0}^{N-1} \sqrt{\mathbb{V}\left[Z_\Delta^l(x,t)\right] C_l})$ ▷ See (7.30)

$N_{l_0}(x,t) \leftarrow \mu(x,t)\sqrt{\mathbb{V}[Z^{l_0}(x,t)]/c_{l_0}}$ ▷ See (7.27)

$M_l(x,t) \leftarrow \mu(x,t)\sqrt{\mathbb{V}[Z_\Delta^l(x,t)]/C_l}$, $l = l_0, \ldots, N-1$ ▷ See (7.28)

$N_{l_0}(t) \leftarrow \text{ceiling}(\max_x N_{l_0}(x,t))$

$M_l(t) \leftarrow \text{ceiling}(\max_x M_l(x,t))$, $l = l_0, \ldots, N-1$

$\Sigma \leftarrow 0$ ▷ Initialize the samples sum for $Z^{l_0}(x,t)$

for $i = 1, \ldots, N_{l_0}(t)$ **do** ▷ Monte Carlo loop for $\mathbb{E}[Z^{l_0}(x,t)]$

  Set a realization of $S_0^{l_0}(t)$ and $S_1^{l_0}(t)$

  if $|S_0^{l_0}(t)| \geq |S_1^{l_0}(t)|$ **then**

    Set a realization of Y_1

    $\Sigma \leftarrow \Sigma + Z_0^{l_0}(x,t)$ ▷ See (7.4)

  else

    Set a realization of Y_0

    $\Sigma \leftarrow \Sigma + Z_1^{l_0}(x,t)$ ▷ See (7.4)

  end if

end for

$\mathbb{E}[Z^{l_0}(x,t)] \leftarrow \Sigma / N_{l_0}(t)$ ▷ Set sample average for $Z^{l_0}(x,t)$

for $l = l_0, \ldots, N-1$ **do** ▷ Loop through levels

  $\Sigma \leftarrow 0$ ▷ Initialize the samples sum for $Z_\Delta^l(x,t)$

  for $i = 1, \ldots, M_l(t)$ **do** ▷ Monte Carlo loop for $\mathbb{E}[Z_\Delta^l(x,t)]$

    Set a realization of $S_0^l(t)$, $S_1^l(t)$, $S_0^{l+1}(t)$ and $S_1^{l+1}(t)$

    if $U_l = 0$ **then** ▷ See (7.22)–(7.23)

      Set a realization of Y_1

      $\Sigma \leftarrow \Sigma + \Delta Z_0^l(x,t)$ ▷ See (7.21)

    else

      Set a realization of Y_0

      $\Sigma \leftarrow \Sigma + \Delta Z_1^l(x,t)$ ▷ See (7.21)

    end if

  end for

  $\mathbb{E}[Z_\Delta^l(x,t)] \leftarrow \Sigma / M_l(t)$ ▷ Set sample average for $Z_\Delta^l(x,t)$

end for

$f_X^{N,M}(x,t) \leftarrow \mathbb{E}[Z^{l_0}(x,t)] + \sum_{l=l_0}^{N-1} \mathbb{E}[Z_\Delta^l(x,t)]$ ▷ See (7.24)

Return $f_X^{N,M}(x,t)$ ▷ Approximation of $f_{X^{N(t)}}(x)$

is made heuristically, depending on the size of N for which (7.34) is greater than 1.

We showcase the multilevel Monte Carlo method in numerical experiments. The first example deals with (3.1) having infinite series expansions in the input data. We will showcase the methodology proposed step-by-step. We will observe that (7.34) is greater than 1 for $N \geq 1$, therefore multilevel Monte Carlo will always show superior to standard Monte Carlo. In the second example, (7.34) will be greater than 1 only for large N. To take advantage of multilevel Monte Carlo, the error δ will be chosen small, so that N will be large enough.

Example 7.13 Consider (3.1) with $t_0 = 0$, $Y_0, Y_1 \sim \text{Normal}(2, 3)$, $B(t) = t$ and

$$A(t) = \sum_{n=0}^{\infty} \frac{U_n}{n!} t^n,$$

where $U_0, U_1, \ldots$ are independent random variables $\text{Beta}(11, 15)$. The series defining $A(t)$ converges in $\mathrm{L}^\infty(\Omega)$ for all $t \in \mathbb{R}$. There is a unique analytic stochastic process $X(t)$ that solves (3.1) in the $\mathrm{L}^p(\Omega)$-sense, $1 \leq p < \infty$.

Algorithm 4 is run for $N \leq 20$, using $m_{\mathbb{V}} = 10,000$ realizations. Using Algorithm 5 (with costs $c_l = C_l = l^2$), we estimate the ratios of complexities (7.33) for $l_1 \in [1, 19]$, $N \in [l_1 + 1, 20]$. In Figure 7.22 we present the ratios. Observe that they are greater than 1, therefore indicating higher efficiency of multilevel Monte Carlo. For each N, we keep the level l_0 maximizing the ratio, see (7.34). For $N \geq 7$, the same maximizer $l_0 = 6$ is obtained.

Figure 7.23 reports the ratios for $N > l_0 = 6$. The ratios grow roughly quadratically with N, with growth constant $1/l_0^2$. This is because the complexity of standard Monte Carlo increases quadratically with N, while the cost of multilevel Monte Carlo becomes constant from a certain level N. Small oscillations may be observed because of the finite number of simulations $m_{\mathbb{V}}$ used to estimate the variances.

Once we have the parameters of the multilevel Monte Carlo strategy, we aim at approximating pointwise the probability density function of $X(t)$ at $t = 1.5$, $f_{X(t=1.5)}(x)$. Figure 7.24 reports in log-scale the estimated bias errors (7.35) with N (the maximum in x of (7.35) has been taken). We fix $N_\infty = 100$ and consider $N \in [1, 20]$. The estimates are computed with standard Monte Carlo simulation, using $20,000$ realizations. Exponential decrease of the bias error is perceived. The oscillating behavior of the bias estimates in the plot is due to

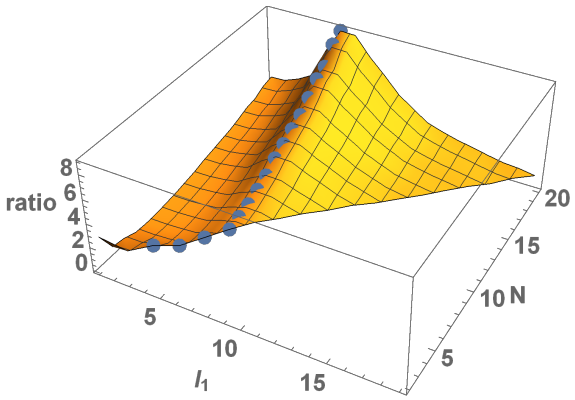

Figure 7.22: Ratios of complexities between standard Monte Carlo and multilevel Monte Carlo (see (7.33)). The points indicate the maximum ratio with l_1 per truncation N. This figure corresponds to Example 7.13.

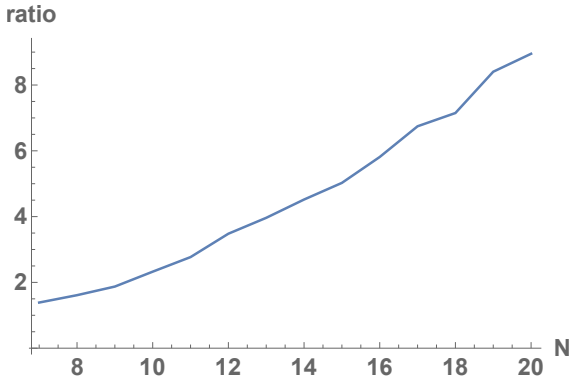

Figure 7.23: Ratios of complexities between standard Monte Carlo and multilevel Monte Carlo (see (7.33)), for $N > l_0 = 6$. This figure corresponds to Example 7.13.

their possible non-monotonic decay, and to the finite number of Monte Carlo realizations.

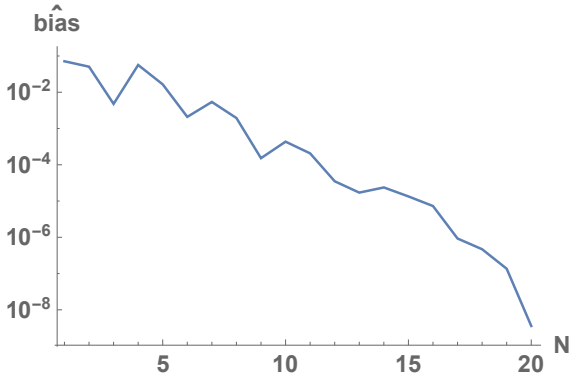

Figure 7.24: Estimated bias errors with N in log-scale. This figure corresponds to Example 7.13.

Suppose that we wish to approximate $f_{X(t=1.5)}(x)$ uniformly with root mean square error $\delta > 0$. Let $\epsilon = \epsilon_{\text{bias}} = \epsilon_{\text{sampling}} = \delta/2$. We pick the first $N = N(\epsilon)$ for which the estimated bias error is less than ϵ. Multilevel Monte Carlo can be applied with $l_0 = 6$, that $N(\epsilon)$, and statistical error variance ϵ^2. In Figure 7.25, we illustrate the relation between ϵ and the complexities of multilevel and standard Monte Carlo, with maximum level $N = N(\epsilon)$, using (7.31)–(7.32). The following approximate relations hold:

$$W^{\text{MC}}_{N(\epsilon)} \propto \epsilon^{-2} N(\epsilon)^2, \quad W^{\text{multi}}_{l_0,N(\epsilon)} \propto \epsilon^{-2} l_0^2 \tag{7.36}$$

(the symbol $\propto$ denotes proportionality to a constant). The cost of multilevel Monte Carlo is reduced by factor $l_0^2/N(\epsilon)^2$. In log-log scale, the complexity of multilevel Monte Carlo increases linearly with respect to ϵ^{-1}, with slope 2. On the contrary, for standard Monte Carlo, such relation is shifted up with $2 \log N(\epsilon)$. The sequence $N(\epsilon)$ increases very slowly with $\epsilon \to 0$, because of the fast decrease of the bias errors towards 0. But as $\epsilon \to 0$, multilevel Monte Carlo becomes more and more efficient compared with standard Monte Carlo.

Table 7.1 reports, for different statistical error variances ϵ^2, the first $N(\epsilon)$ for which the estimated bias error is less than ϵ, the complexities of the multilevel and the standard Monte Carlo strategies using (7.31)–(7.32), and the number of simulations. The indices $N(\epsilon)$ grow slowly with ϵ, because of the fast decay of the bias errors to 0. The cost of the multilevel approach is smaller than

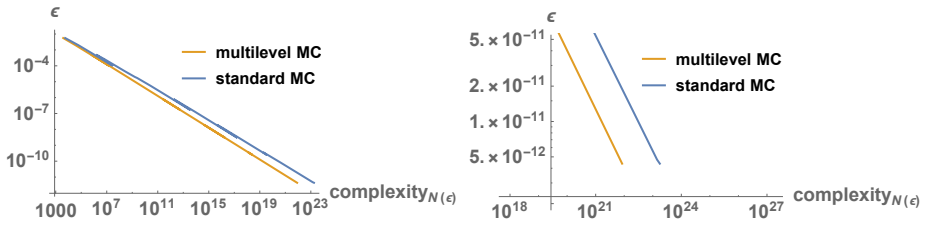

Figure 7.25: Sampling error ϵ versus complexity using (7.31)–(7.32) (with $N(\epsilon)$), in log-log scale. The last plot is a zoom. This figure corresponds to Example 7.13.

ϵ	$N(\epsilon)$	$W^{\text{multi}}_{l_0,N(\epsilon)}$	$N_{l_0}; M_{l_0}, \ldots, M_{N(\epsilon)-1}$	$W^{\text{MC}}_{N(\epsilon)}$	samples MC
0.002	8	3.5×10^4	1173; 107, 52	5.8×10^4	913
0.001	9	1.5×10^5	4877; 447, 216, 99	3.0×10^5	3673
0.0005	9	6.1×10^5	19507; 1787, 861, 395	1.2×10^6	14692
0.00025	9	2.4×10^6	78028; 7147, 3442, 1580	4.8×10^6	58767
0.0001	9	1.5×10^7	487670; 44669, 21508, 9873	3.0×10^7	367291
0.00005	12	6.4×10^7	1974873;181699,87707,42121,12293,5128,2859	2.2×10^8	1507941

Table 7.1: Sampling error ϵ, index $N(\epsilon)$, complexities of the multilevel and the standard Monte Carlo strategies using (7.31)–(7.32), and the number of simulations. This table corresponds to Example 7.13.

standard Monte Carlo; for $N = 9$ it is half of the cost, for $N = 12$ it is decreased by factor greater than three (see also Figure 7.23 concerning the ratios of the complexities). In the multilevel Monte Carlo method, most of the simulations are localized at the first level $l_0 = 6$ with N_{l_0} realizations, while for the increments the number of realizations M_l is decreasing.

The sampling errors ϵ considered in Figure 7.25 and Table 7.1 are theoretical. Given a theoretical root mean square error δ, $\epsilon = \delta/2$ and $N = N(\epsilon)$, the actual root mean square error $\hat{\delta} = \max_x \|f_{X(t)}(x) - \hat{f}_{X^N(t)}(x)\|_2$ is smaller than δ. This is presented in Figure 7.26. As the exact density function $f_{X(t=1.5)}(x)$ is unknown, we first estimate it with very small bias and sampling errors, and regard this estimation as the exact density function. Then the real error $\hat{\delta}$ is estimated via a sampled root mean square error from several realizations of $\hat{f}_{X^N(t=1.5)}$. In Figure 7.26, we observe that the multilevel strategy is more efficient than the standard Monte Carlo simulation, as expected from our theoretical analysis in Figure 7.25.

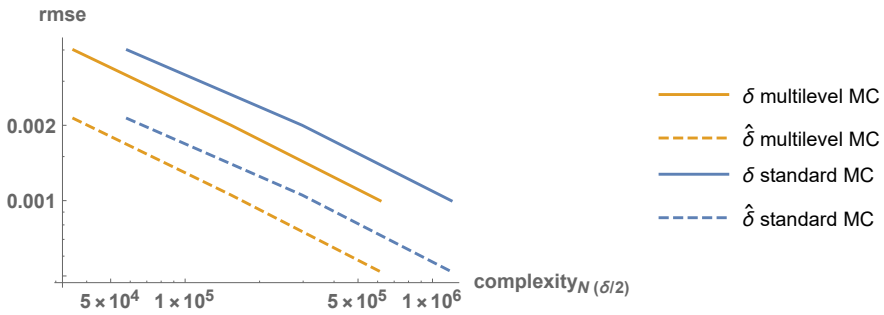

Figure 7.26: Root mean square error (rmse) versus complexity using (7.31)–(7.32) (with $N(\delta/2)$), in log-log scale. We plot the theoretical rmse δ (solid lines) and the real rmse $\hat{\delta}$ (dashed lines). This figure corresponds to Example 7.13.

Let us approximate $f_{X(t=1.5)}(x)$ uniformly with theoretical root mean square error $\delta = 0.001$. Let $\epsilon = \delta/2 = 0.0005$. We pick the first N for which the estimated bias error is less than ϵ: $N = N(\epsilon) = 9$. Then we apply the multilevel Monte Carlo procedure from Algorithm 6 with statistical error variance ϵ^2. The optimal number of simulations for $N = 9$ is the following (see Table 7.1): $M = (N_6; M_6, M_7, M_8) = (19507; 1787, 861, 395)$. As expected, most of the cost is localized at level 6, while for the increments the cost goes down. The complexity of multilevel Monte Carlo is 6.1×10^5, which is smaller than that of standard Monte Carlo, 1.2×10^6. In Figure 7.27, first panel, we plot the estimate $\hat{f}_{X^{N=9}(t=1.5)}(x) = f_X^{N=9,M}(x, t = 1.5)$ (output of Algorithm 6). It satisfies $\|f_{X(1.5)}(x) - \hat{f}_{X^9(1.5)}(x)\|_2 \leq \delta$, for all $x \in \mathbb{R}$. In the second plot, we show the estimated densities for $N = 8$, 9 and 10 computed with Algorithm 6, with statistical error standard deviation $\epsilon = 0.0005$. Observe that the densities overlap, thus indicating the expected pointwise convergence.

Example 7.14 We study problem (7.11) with one degree polynomial coefficients. The probability distributions for the random inputs are $A_0 = 4$, $A_1 \sim \text{Uniform}(-3, 1)$, $B_0 \sim \text{Gamma}(2, 2)|_{[0,4]}$, $B_1 \sim \text{Bernoulli}(0.35)$ and $Y_0, Y_1 \sim \text{Normal}(2, \sigma = 3)$, all them independent. Due to the boundedness of A_0, A_1, B_0 and B_1, this problem has a unique analytic stochastic solution $X(t)$, $t \in \mathbb{R}$, in the sense of the $\text{L}^p(\Omega)$ random calculus, for $1 \leq p < \infty$.

The objective is to approximate the probability density function $f_{X(t)}(x)$ using the multilevel Monte Carlo approach, by estimating $f_{X^N(t)}(x)$. The theoretical

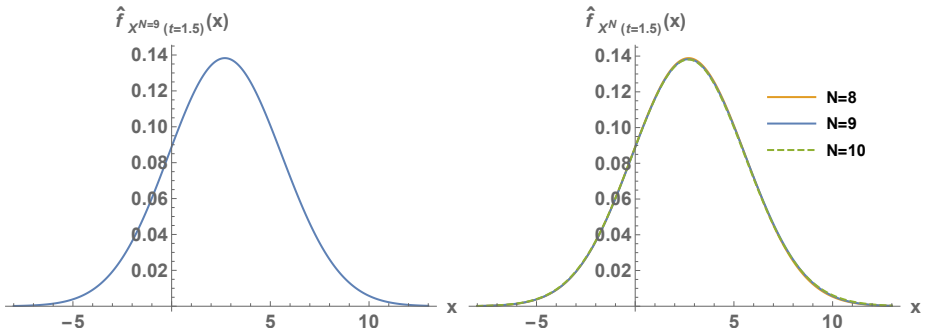

Figure 7.27: Estimate $\hat{f}_{X^9(1.5)}(x)$. Estimates $\hat{f}_{X^N(1.5)}(x)$, for $N = 8, 9, 10$. This figure corresponds to Example 7.13.

results guarantee that $\{f_{X^N(t)}(x)\}_{N\geq 0}$ converges to $f_{X(t)}(x)$ pointwise and in $\mathrm{L}^p(\mathbb{R})$, $1 \leq p < \infty$, as $N \to \infty$. We work at time $t = 1.5$.

We run Algorithm 4 for N large (till $N = 70$), using parallel computing and $m_{\mathbb{V}} = 50,000$ realizations. The number of realizations $m_{\mathbb{V}}$ is actually excessively large, but we want to be quite exact in this example to show accurate complexities.

Using Algorithm 5 (with costs $c_l = C_l = l$), we obtain the ratios of complexities between standard Monte Carlo and multilevel Monte Carlo, see (7.33). In Figure 7.28, we vary the starting level $l_1 \in [1, 69]$ and the maximum level $N \in [l_1 + 1, 70]$ and plot the corresponding ratios. Observe that the maximums (7.34) are greater than 1 (indicating that multilevel Monte Carlo has higher efficiency) for $N \geq 36$ only, in contrast to Example 7.13. For each N, we keep l_0 maximizing the ratio, see (7.34). For $N = 36$, the maximizer is $l_0 = 35$; for $N = 37$ and 38, the maximizer is $l_0 = 36$. For all $N \geq 39$, it is obtained the same $l_0 = 38$ as the maximizer.

In Figure 7.29 we report the ratios for $N > l_0 = 38$. The ratios are greater than 1 and, as N increases, they tend to infinity at linear rate approximately, although with small slope $\approx 1/l_0$. The complexity of standard Monte Carlo grows linearly to infinity with N, while the complexity of multilevel Monte Carlo becomes constant from a certain N. The oscillating behavior of the depicted ratios is due to the statistical error driven by the previous Monte Carlo estimates for the variances. When $m_{\mathbb{V}}$ gets larger, the oscillations are mitigated and more exact results are obtained.

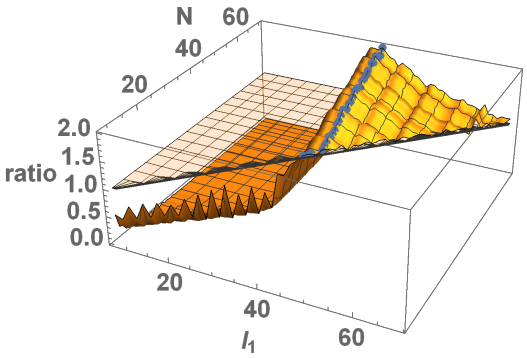

Figure 7.28: Ratios of complexities between standard Monte Carlo and multilevel Monte Carlo (see (7.33)). The points indicate the maximum ratio with l_1 per truncation N. The transparent horizontal surface represents the ratio 1. This figure corresponds to Example 7.14.

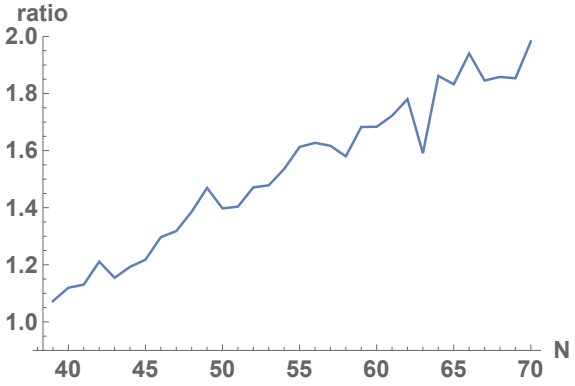

Figure 7.29: Ratios of complexities between standard Monte Carlo and multilevel Monte Carlo (see (7.33)), for $N > l_0 = 38$. This figure corresponds to Example 7.14.

Let us approximate $f_{X(t=1.5)}(x)$ uniformly with theoretical root mean square error $\delta = 0.0001$. Figure 7.30 shows the estimated bias errors (7.35), with $N_\infty = 200$ and 20,000 Monte Carlo simulations (the maximum in x of (7.35) has been taken). We denote the bias errors as ϵ^N_{bias}. The bias error needs quite large truncation order N to start decreasing (up to $N = 25$ the bias is 0.1 approximately), and for $N \geq 35$ it drops abruptly. The estimated bias error is less than δ only for $N \geq 37$, therefore we are in the region where multilevel Monte Carlo is more efficient than standard Monte Carlo. Notice that, in this example, we have fixed δ small on purpose to take advantage of the multilevel Monte Carlo strategy. In the previous Example 7.13, we did not worry about δ nor its decomposition because multilevel Monte Carlo showed more efficient than standard Monte Carlo from the very first truncation order.

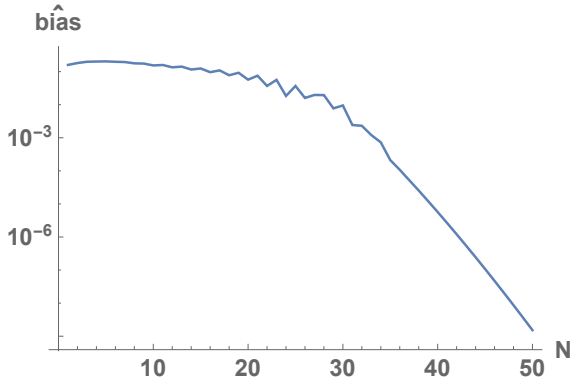

Figure 7.30: Estimated bias errors with N in log-scale. This figure corresponds to Example 7.14.

In Figure 7.31 we plot, with $N \geq 37$, the complexities of the multilevel and the standard Monte Carlo approaches with $\epsilon_{\text{sampling}} = \delta - \epsilon^N_{\text{bias}}$ ($\delta = 0.0001$), using (7.31)–(7.32). We observe that the complexity of multilevel Monte Carlo is smaller, as expected. With formulas, we have

$$W^{\text{MC}}_N(\epsilon_{\text{sampling}}) \propto (\delta - \epsilon^N_{\text{bias}})^{-2} N, \quad W^{\text{multi}}_{l_0,N}(\epsilon_{\text{sampling}}) \propto (\delta - \epsilon^N_{\text{bias}})^{-2} l_0.$$

When $N \approx l_0$, we have $\delta - \epsilon^N_{\text{bias}} \approx 0$, so the complexities are large. When N is large, then $\delta - \epsilon_{\text{bias}} \approx \delta$, so $W^{\text{MC}}_N(\epsilon_{\text{sampling}}) \propto N \nearrow \infty$ and $W^{\text{multi}}_{l_0,N}(\epsilon_{\text{sampling}}) \searrow$ constant.

Given $\delta = 0.0001$, in the case of standard Monte Carlo the minimum complexity is attained when $N = 43$ (bias $\epsilon^{N=43}_{\text{bias}} = 5 \times 10^{-7}$), this being 2×10^8.

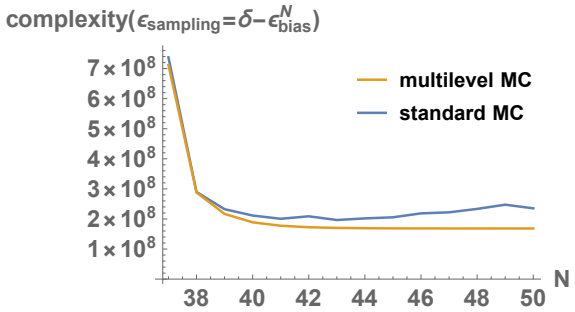

Figure 7.31: Complexities of the standard and multilevel Monte Carlo approaches, see (7.31)–(7.32), with $\epsilon_{\text{sampling}} = \delta - \epsilon_{\text{bias}}^{N}$, $\delta = 0.0001$ fixed. This figure corresponds to Example 7.14.

For $N = 50 > l_0 = 38$, we run the multilevel Monte Carlo algorithm with statistical error variance $\epsilon_{\text{sampling}}^2 = (\delta - \epsilon_{\text{bias}}^{N=50})^2$. We choose $N = 50$ because the corresponding cost of multilevel Monte Carlo is uniformly cheaper than standard Monte Carlo applied with any order of truncation. Also, the complexity of multilevel Monte Carlo becomes constant for $N \geq 50$. For $N = 50$, the optimal number of simulations is $M = (N_{38}; M_{38}, \ldots, M_{50}) = (4527935; 12447, 4977, 2624, 1507, 475, 292, 126, 45, 22, 9, 4, 2)$, with a complexity that is 1.7×10^8. Figure 7.32, first panel, plots the graph of the estimated density function $\hat{f}_{X^{N=50}(t=1.5)}(x)$, which approximates the true density $f_{X(t=1.5)}(x)$ with root mean square error less than δ, for all $x \in \mathbb{R}$. The second plot depicts the density estimates for $N = 49$, 50 and 51, showing the expected overlapping.

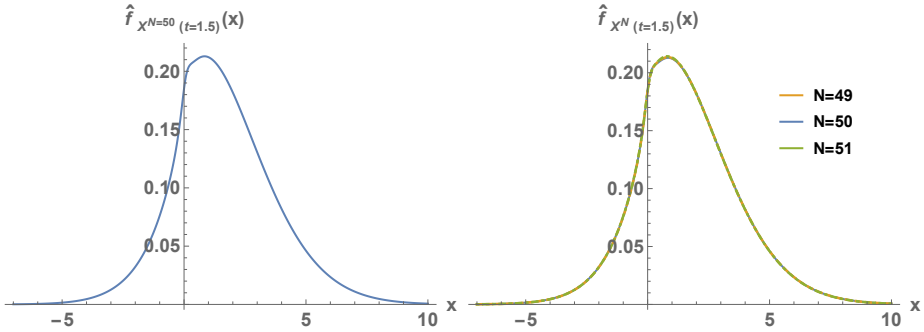

Figure 7.32: Estimate $\hat{f}_{X^{50}(1.5)}(x)$. Estimates $\hat{f}_{X^{N}(1.5)}(x)$, for $N = 49, 50, 51$. This figure corresponds to Example 7.14.

Remark 7.15 Our application of the multilevel Monte Carlo strategy is based on $\mathrm{L}^\infty(\mathbb{R})$ norms. First, we acted pointwise in x in the constrained minimization problem from Proposition 7.12, and in Algorithm 6 we took the $\mathrm{L}^\infty(\mathbb{R})$ norm of the optimal number of simulations. Then the target density function was approximated uniformly in x. Second, the ratio (7.33) of complexities considered the $\mathrm{L}^\infty(\mathbb{R})$ norms of the standard and the multilevel Monte Carlo approaches.

Other norms can be considered. For instance, we could minimize the cost (7.26) given a fixed $\mathrm{L}^1(\mathbb{R})$ norm of the statistical error variance (7.25): $\int_{\mathbb{R}} \| f_{X^N(t)}(x) - f_X^{N,M}(x,t) \|_2^2 \, \mathrm{d}x = \epsilon^2$. In this case, we substitute each variance in x by its integral on $\mathbb{R}$. Then the approach becomes independent of x from the beginning.

When the cost is minimized pointwise in x, another possibility consists in considering the ratio of the $\mathrm{L}^1(\mathbb{R})$ norms of the complexities:

$$W_N^{\mathrm{MC},\|\cdot\|_1}(t) = \epsilon^{-2} c_N \int_{\mathbb{R}} \mathbb{V}[Z^N(x,t)] \, \mathrm{d}x, \tag{7.37}$$

$$W_{l_1,N}^{\mathrm{multi},\|\cdot\|_1}(t) = \epsilon^{-2} \int_{\mathbb{R}} \left(\sqrt{\mathbb{V}\left[Z^{l_1}(x,t)\right] c_{l_1}} + \sum_{l=l_1}^{N-1} \sqrt{\mathbb{V}\left[Z_\Delta^l(x,t)\right] C_l} \right)^2 \mathrm{d}x, \tag{7.38}$$

$$R_{l_1,N}^{\|\cdot\|_1}(t) = \frac{W_N^{\mathrm{MC},\|\cdot\|_1}(t)}{W_{l_1,N}^{\mathrm{multi},\|\cdot\|_1}(t)} \tag{7.39}$$

(the ratio is independent of ϵ). In the two examples from this section, the conclusions derived are analogous considering $\mathrm{L}^1(\mathbb{R})$ norms, thus showing robustness.

We present results regarding Example 7.13. Using $m_{\mathbb{V}} = 10,000$ realizations and costs $c_l = C_l = l^2$, we estimate the ratios of complexities (7.39) for $l_1 \in [1, 19]$ and $N \in [l_1 + 1, 20]$. Figure 7.33, first panel, shows the ratios. These ratios show a similar pattern to Figure 7.22. For each N, we keep the level l_0 maximizing the ratio (7.39). For $N \geq 7$, the same maximizer $l_0 = 6$ is obtained, as in Example 7.13. The second panel of Figure 7.33 compares the two ratios (7.33) and (7.39) for $N \geq l_0 = 6$. The two ratios are very similar, only showing a small discrepancy in part due to the Monte Carlo procedure with $m_{\mathbb{V}}$ realizations. The fact that the two ratios are practically coinciding is justified theoretically by the approximation $R_{l_1,N}^{\|\cdot\|_1}(t) \approx N^2/l_1^2$, which also held for (7.33).

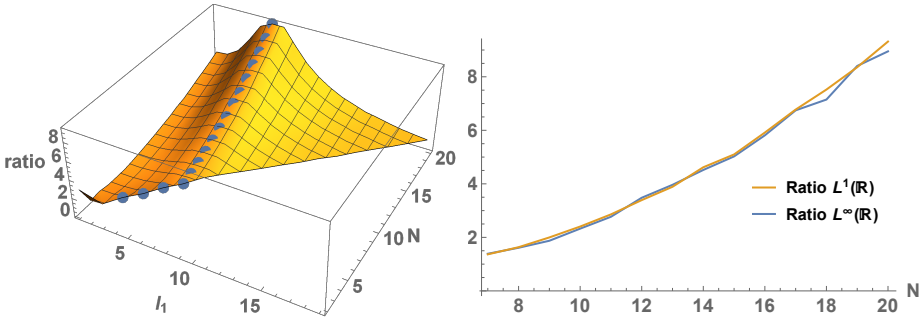

Figure 7.33: First panel: ratios of the $\mathrm{L}^1(\mathbb{R})$ norms of the complexities between standard and multilevel Monte Carlo (see (7.39)), where the points indicate the maximum ratio with l_1 per truncation N. Second panel: ratios of complexities between standard and multilevel Monte Carlo with $\mathrm{L}^\infty(\mathbb{R})$ norms (see (7.33)) and $\mathrm{L}^1(\mathbb{R})$ norms (see (7.39)), for $N > l_0 = 6$. This figure corresponds to Remark 7.15.

Figure 7.34 analyzes theoretical error against complexity, using (7.37)–(7.38). It corresponds to Figure 7.25 by considering $\mathrm{L}^1(\mathbb{R})$ norms, instead. When changing the norm, the "line" corresponding to error versus complexity in log-log scale is shifted up or down but has the same slope. This is because changing the norm only modifies the proportionality constants in (7.36). This is illustrated in Figure 7.35, where the two "lines" corresponding to multilevel Monte Carlo are parallel, and also for standard Monte Carlo.

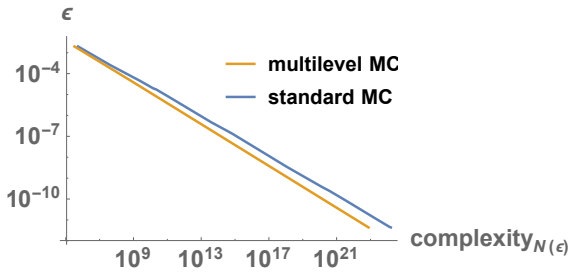

Figure 7.34: Sampling error ϵ versus $\mathrm{L}^1(\mathbb{R})$ norm of the complexity (with $N(\epsilon)$), see (7.37)–(7.38), in log-log scale. This figure corresponds to Remark 7.15.

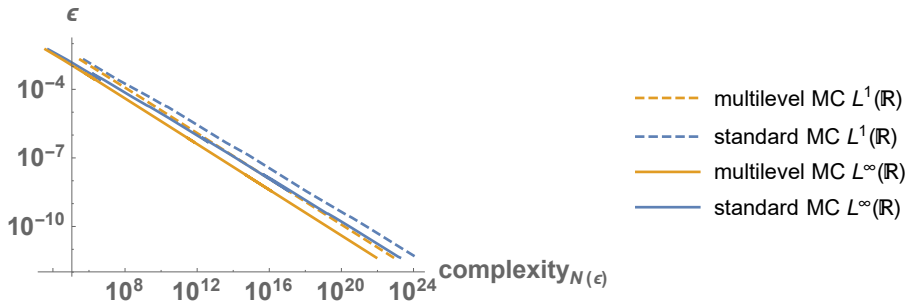

Figure 7.35: Sampling error ϵ versus the $\mathrm{L}^1(\mathbb{R})$ (see (7.37)–(7.38)) and $\mathrm{L}^\infty(\mathbb{R})$ (see (7.31)–(7.32)) norms of the complexity (with $N(\epsilon)$), in log-log scale. This figure corresponds to Remark 7.15.

7.5 Conclusions

This chapter focused on the improvement of the computational methods and algorithms from Chapter 6. In Chapter 6, we studied the random second-order linear differential equation with analytic data processes and random initial conditions, and constructed approximations to the probability density function of the stochastic solution based on dimensionality reduction and standard Monte Carlo sampling strategy. The main goal in this chapter is to speed up the convergence of the standard Monte Carlo algorithm used in Chapter 6 by controlling the variance of the statistical estimators, thus avoiding noisy features in the density estimates.

The path-wise selection of the initial condition used in the density expression leads to a stable algorithm with good convergence. It was proved theoretically that the statistical estimators constructed with this method always possess finite variance. The implementation of the algorithm was illustrated through simple autonomous and non-autonomous linear differential equations.

Improvements are achieved by computing the expectation with respect to the initial conditions via numerical integration methods. By considering the densities of the initial conditions as integration weights, we focused on Gaussian quadrature rules. The orthogonal polynomials were considered with respect to standard probability distributions. Several conclusions are reached from this approach. When the initial conditions are comparable with the input coefficients of the equation in terms of random variability, the quadrature rule for integration allows significant variance reduction of the statistical estima-

tors. Even when the initial conditions have small dispersion compared to the input coefficients, if the cost of each realization of the fundamental set is quite expensive, the quadrature rule yields similar expense to a plain Monte Carlo strategy. The quadrature degree has to be selected sufficiently large to assure error-free approximations of the integrals.

A multilevel version of the Monte Carlo algorithm allows reduction of the complexity, by linear or quadratic factor. The multilevel Monte Carlo approach turns out to be useful when high accuracy in the density approximation is required. The starting level was selected on the fly to maximize the ratio of complexities between the standard and the multilevel Monte Carlo approach. Once the initial level is selected, and given a statistical error variance, we obtained the optimal number of simulations per level to minimize the complexity. The numerical results illustrated the methodology step-by-step in order to show the improvement with regard to the standard version of the Monte Carlo simulation algorithm.

Several fundamental extensions and algorithmic improvements may be conceived, especially concerning the multilevel Monte Carlo methodology. Firstly, the three parts in which we divided the multilevel procedure (variance of the increments, selection of the initial level by maximizing the ratio, and approximation of the density) could be combined in an automated algorithm, which would obtain the bias error, the sampling error, the variances, the initial level and the optimal number of simulations on the fly. Secondly, the multilevel Monte Carlo approach may be combined with a numerical integration scheme for the expectation with respect to the initial conditions. Finally, our multilevel strategy relied on considering consecutive levels, while other strategies could utilize different sequences of levels, such as linear or geometric ones.

An interesting area to explore is the extension of the presently methodology to linear systems of second-order random differential equations. In this case, the aim is at approximating the joint probability density function of the vector-valued stochastic solution. With the exception of simple autonomous linear systems whose density function can be computed through transformations of random variables/vectors and direct integration, no analysis has been carried out for non-autonomous systems. The application to non-autonomous systems would offer promising advances in the active field of random differential equations.

Acknowledgements

This work has been supported by the Spanish "Ministerio de Economía y Competitividad" grant MTM2017–89664–P.

I acknowledge the doctorate scholarship granted by "Programa de Ayudas de Investigación y Desarrollo" (PAID), as well as "Ayudas para movilidad de estudiantes de doctorado de la Universitat Politècnica de València para estancias en 2019", Universitat Politècnica de València, for financing my research stay of three months at Centre de Mathématiques Appliquées (CMAP UMR7641), École Polytechnique, Palaiseau, France.

I am also grateful to Inria (Centre de Saclay and DeFi Team), which hosted me during my research stay at CMAP.

The main results of this chapter have been submitted for publication. This chapter is an international collaboration with the researcher Olivier P. Le Maître (CMAP, CNRS, INRIA, École Polytechnique, Institut Polytechnique de Paris, 91128 Palaiseau, France).

Chapter 8

On the Legendre differential equation with uncertainties at the regular-singular point 1

In this chapter, we construct two linearly independent response processes to the random Legendre differential equation on $(-1,1) \cup (1,3)$*, consisting of* $\mathrm{L}^p(\Omega)$ *convergent random power series around the regular-singular point* 1*. A theorem on the existence and uniqueness of* $\mathrm{L}^p(\Omega)$ *solution to the random Legendre differential equation on the intervals* $(-1,1)$ *and* $(1,3)$ *is obtained. The hypotheses assumed are simple: initial conditions in* $\mathrm{L}^p(\Omega)$ *and random input* A *in* $\mathrm{L}^\infty(\Omega)$ *(this is equivalent to* A *having absolute moments that grow at most exponentially). Thus, this chapter extends the deterministic theory to a random framework. Uncertainty quantification for the solution stochastic process is performed by truncating the random series and taking limits in* $\mathrm{L}^p(\Omega)$*. In the numerical experiments, we approximate its expectation and variance for certain forms of the differential equation. The reliability of our approach is compared with Monte Carlo simulation and gPC expansions.*

8.1 Introduction

The random Legendre differential equation (8.1),

$$\begin{cases} (1-t^2)\ddot{X}(t) - 2t\dot{X}(t) + A(A+1)X(t) = 0, \\ X(t_0) = Y_0, \\ \dot{X}(t_0) = Y_1, \end{cases} \tag{8.1}$$

where A, Y_0 and Y_1 are random variables in a common underlying complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$, has been already studied at the regular point $t = 0$ with initial condition at $t_0 = 0$, see Chapter 2. In [31], a mean square power series solution to (8.1) was constructed on $(-1/\mathrm{e}, 1/\mathrm{e})$, being e the Euler constant. This result has been extended in the recent contribution [21] (it is Chapter 2), where the solution has been constructed on the whole domain $(-1, 1)$ with weaker assumptions on the random input coefficients. A common hypotheses in both works [31, 21] was that the random variable A has statistical absolute moments that increase at most exponentially, equivalently, that A is a bounded random variable, see [21, Lemma 2.2] (Lemma 2.2). This assumption of boundedness for A will be essential in our subsequent development.

The aim of this chapter is to continue extending the classical deterministic results for the Legendre differential equation to the random setting by taking advantage of the so-called $\mathrm{L}^p(\Omega)$, $1 \leq p \leq \infty$, random calculus. It is worth pointing out that the $\mathrm{L}^p(\Omega)$ random calculus has been widely used to study both theoretical and numerically random differential equations [77, 115, 121, 127, 168].

Our goal is to construct a fundamental set around the regular-singular point $t = 1$, as well as to quantify reliable approximations to the main statistical functions of the solution stochastic process to the initial value problem (8.1). To the best of our knowledge, this is the first contribution in the extant literature where the regular-singular point case is addressed for a random second-order differential equation. In this sense, we want to point out that the subsequent approach may be useful to study other important random second-order linear differential equations around regular-singular points. This kind of differential equations are met in many physical and engineering problems [106, 110, 133]. In particular, the Legendre differential equation is very useful for treating the boundary value problems exhibiting spherical symmetry.

The organization of the present chapter is as follows. In Section 8.2, the main results regarding the random initial value problem (8.1) are stated and

proved. In Section 8.3, we show how to approximate the moments of the solution stochastic process by truncating the corresponding random power series. Section 8.4 is devoted to illustrate our theoretical findings via different examples of the random initial value problem (8.1), where approximations of the expectation and the variance functions of the solution stochastic process are computed and compared with Monte Carlo simulation and gPC expansions. Finally, conclusions are drawn in Section 8.5.

8.2 Random Legendre differential equation at the regular singular point 1

As it has been previously indicated, our goal in this section is to provide an analogous analysis to [21, 31] for the regular-singular point $t = 1$ ([21] corresponds to Chapter 2).

By the deterministic theory on the Legendre differential equation, if A is constant, then a deterministic fundamental set $\{\phi_1(t), \phi_2(t)\}$ is given by

$$\phi_1(t) = \sum_{n=0}^{\infty} c_n (t-1)^n, \quad |t-1| < 2,$$

where $\{c_n\}_{n=0}^{\infty}$ is defined by the recursive relation

$$c_0 = 1, \quad c_{n+1} = \frac{(n+1-A)(A-n)}{2(n+1)^2} c_n, \quad n = 0, 1, 2, \ldots,$$

and

$$\phi_2(t) = \phi_1(t) \log|t-1| + \sum_{n=1}^{\infty} d_n (t-1)^n, \quad |t-1| < 2,$$

where $\{d_n\}_{n=1}^{\infty}$ is defined as follows:

$$d_1 = \frac{-c_0 - 4c_1}{2}, \; d_{n+1} = -\frac{(n+1+A)(n-A)d_n + 4(n+1)c_{n+1} + (2n+1)c_n}{2(n+1)^2}.$$

This may be checked by direct calculation, see [33, p. 205, pp. 217–218].

In our setting, we first randomize this fundamental set of solutions. Therefore, we consider the following two stochastic processes

$$X_1(t) = \sum_{n=0}^{\infty} C_n (t-1)^n, \quad X_2(t) = X_1(t) \log|t-1| + \sum_{n=1}^{\infty} D_n (t-1)^n, \quad (8.2)$$

where the coefficients $\{C_n\}_{n=0}^{\infty}$ and $\{D_n\}_{n=1}^{\infty}$ are random variables defined in our complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ recursively as follows:

$$C_0 = 1, \quad C_{n+1} = \frac{(n+1-A)(A-n)}{2(n+1)^2} C_n, \quad n = 0, 1, 2, \ldots, \tag{8.3}$$

and

$$D_1 = \frac{-C_0 - 4C_1}{2}, \; D_{n+1} = -\frac{(n+1+A)(n-A)D_n + 4(n+1)C_{n+1} + (2n+1)C_n}{2(n+1)^2}. \tag{8.4}$$

In the following Proposition 8.1, we will see that both series (8.2) converge in $\mathrm{L}^{\infty}(\Omega)$ on $(-1,1) \cup (1,3)$ (i.e., radius of convergence 2, so retaining their deterministic counterpart), in particular, pointwise on $\omega \in \Omega$ for $t \in (-1,1) \cup (1,3)$. This $\mathrm{L}^{\infty}(\Omega)$ convergence will imply that the stochastic processes $X_1(t)$ and $X_2(t)$ are solutions to the random Legendre differential equation on the domain $(-1,1) \cup (1,3)$ in the $\mathrm{L}^{\infty}(\Omega)$ sense, since the random power series can be differentiated in the $\mathrm{L}^{\infty}(\Omega)$ sense term by term.

Proposition 8.1 *If $A \in \mathrm{L}^{\infty}(\Omega)$, then both series defined by (8.2)–(8.4) converge in $\mathrm{L}^{\infty}(\Omega)$ on the interval $(-1,1) \cup (1,3)$. In particular, the stochastic processes $X_1(t)$ and $X_2(t)$, given by (8.2)–(8.4), are solutions to the random Legendre differential equation on the domain $(-1,1) \cup (1,3)$ in the $\mathrm{L}^{\infty}(\Omega)$ sense.*

Proof. First, let us prove that the random power series defining $X_1(t)$ converges in $\mathrm{L}^{\infty}(\Omega)$, for every $t \in (-1,1) \cup (1,3)$. From (8.3), we have

$$\|C_n\|_{\infty} \leq \frac{(n + \|A\|_{\infty})(n + \|A\|_{\infty} + 1)}{2n^2} \|C_{n-1}\|_{\infty},$$

that is,

$$\frac{\|C_n\|_{\infty}}{\|C_{n-1}\|_{\infty}} \leq \frac{(n + \|A\|_{\infty})(n + \|A\|_{\infty} + 1)}{2n^2}.$$

Now if we multiply both sides of this last inequality by r, being $0 < r < 2$, and afterwards we take limits as $n \to \infty$, then we obtain that

$$\limsup_{n\to\infty} \frac{\|C_n\|_{\infty} r^n}{\|C_{n-1}\|_{\infty} r^{n-1}} \leq \lim_{n\to\infty} r \frac{(n + \|A\|_{\infty})(n + \|A\|_{\infty} + 1)}{2n^2} = \frac{r}{2} < 1.$$

By applying the d'Alembert's ratio test for numerical series, we derive that the series with general term $\|C_n\|_{\infty} r^n$ is convergent, i.e.,

$$\sum_{n=0}^{\infty} \|C_n\|_{\infty} r^n < \infty,$$

for $0 < r < 2$. This implies that the random power series $X_1(t)$, given by (8.2), has radius of convergence 2 in the Banach space $(\mathrm{L}^\infty(\Omega), \|\cdot\|_\infty)$.
Now let us check that

$$\sum_{n=1}^{\infty} \|D_n\|_\infty r^n < \infty,$$

for $0 < r < 2$. Since $\sum_{n=0}^{\infty} \|C_n\|_\infty r^n < \infty$, the sequence $\{\|C_n\|_\infty r^n\}_{n=0}^{\infty}$ is bounded by a number $M > 0$. That is, $\|C_n\|_\infty \leq M/r^n$, $n \geq 0$. Using this inequality in (8.4), we obtain

$$\begin{aligned}\|D_{n+1}\|_\infty &\leq \frac{(n+\|A\|_\infty)(n+\|A\|_\infty+1)\|D_n\|_\infty + (2n+1)\frac{M}{r^n} + 4(n+1)\frac{M}{r^{n+1}}}{2(n+1)^2} \\ &\leq \frac{(n+\|A\|_\infty+1)^2\|D_n\|_\infty + (2n+1)\frac{M}{r^n} + 4(n+1)\frac{M}{r^{n+1}}}{2(n+1)^2}. \end{aligned} \tag{8.5}$$

Let us define the following sequence of positive numbers $\{H_n : n \geq 1\}$:

$$H_1 = \|D_1\|_\infty,$$

$$H_{n+1} = \frac{(n+\|A\|_\infty+1)^2 H_n + (2n+1)\frac{M}{r^n} + 4(n+1)\frac{M}{r^{n+1}}}{2(n+1)^2}, \quad n = 1, 2, \ldots.$$

From inequality (8.5), it is evident that the sequence $\{H_n : n \geq 1\}$ majorizes $\{\|D_n\|_\infty : n \geq 1\}$, that is,

$$\|D_n\|_\infty \leq H_n, \quad n \geq 1.$$

Now, let us define

$$K_n = \max_{1 \leq k \leq n} H_k r^k.$$

We obtain

$$\begin{aligned} H_{n+1} r^{n+1} &\leq \frac{r(n+\|A\|_\infty+1)^2}{2(n+1)^2} H_n r^n + \frac{r(2n+1)M + 4(n+1)M}{2(n+1)^2} \\ &\leq \frac{r(n+\|A\|_\infty+1)^2}{2(n+1)^2} K_n + \frac{r(2n+1)M + 4(n+1)M}{2(n+1)^2}. \end{aligned}$$

Observe

$$\lim_{n\to\infty} \frac{r(n+\|A\|_\infty+1)^2}{2(n+1)^2} = \frac{r}{2} < 1, \quad \lim_{n\to\infty} \frac{r(2n+1)M + 4(n+1)M}{2(n+1)^2} = 0. \tag{8.6}$$

Then, for $0 < \epsilon < 1 - r/2$ arbitrary but fixed, using the limits in (8.6), we can choose $n_0 = n_0(r, \|A\|_\infty, M)$ such that, for all $n \geq n_0$,

$$\frac{r(n+\|A\|_\infty+1)^2}{2(n+1)^2} < 1 - \epsilon, \quad \frac{r(2n+1)M + 4(n+1)M}{2(n+1)^2} < 1.$$

Thus, for $n \geq n_0$,

$$H_{n+1} r^{n+1} \leq (1-\epsilon) K_n + 1.$$

Suppose that $(1-\epsilon)K_n + 1 \leq K_n$, for all $n \geq n_0$. This implies

$$H_{n+1} r^{n+1} \leq K_n,$$

so that $K_{n+1} = K_n$ for $n \geq n_0$. Let $K = K_n$, $n \geq n_0$. Then

$$H_n \leq K/r^n, \quad n \geq n_0,$$

therefore

$$\sum_{n=n_0}^{\infty} H_n r_0^n \leq K \sum_{n=n_0}^{\infty} H_n (r_0/r)^n < \infty,$$

for each $0 < r_0 < r$. As $0 < r < 2$ is arbitrary, we conclude that

$$\sum_{n=n_0}^{\infty} \|D_n\|_\infty r^n \leq \sum_{n=n_0}^{\infty} H_n r^n < \infty,$$

as wanted.

Otherwise, if there is a strictly increasing sequence of natural numbers $\{n_l\}_{l=1}^{\infty}$ such that $(1-\epsilon)K_{n_l} + 1 > K_{n_l}$, for all $l \geq 1$, we arrive at $K_{n_l} < 1/\epsilon$, $l \geq 1$. Since the sequence $\{K_n\}_{n=1}^{\infty}$ is increasing, we deduce that $K_n < 1/\epsilon$, for all $n \geq 1$. Let $K = 1/\epsilon$, so that

$$H_n \leq K/r^n, \quad n \geq 1.$$

The same reasoning as in the previous paragraph applies in this case, and we are done.

□

Theorem 8.2 *Let $1 \leq p \leq \infty$ and $t_0 \in I$, where I is either $(-1,1)$ or $(1,3)$. Given two initial conditions $X(t_0) = Y_0$ and $\dot{X}(t_0) = Y_1$ that belong to $\mathrm{L}^p(\Omega)$ and if $A \in \mathrm{L}^\infty(\Omega)$, then there exists a unique response process $X(t)$ in the $\mathrm{L}^p(\Omega)$ sense to (8.1) on I. This solution process $X(t)$ has the form*

$$X(t) = A_1 X_1(t) + A_2 X_2(t), \tag{8.7}$$

where

$$A_1 = \frac{Y_0 \dot{X}_2(t_0) - Y_1 X_2(t_0)}{W(X_1, X_2)(t_0)}, \quad A_2 = \frac{Y_1 X_1(t_0) - Y_0 \dot{X}_1(t_0)}{W(X_1, X_2)(t_0)}, \tag{8.8}$$

and $W(X_1, X_2)(t_0)$ is the Wronskian of the pair $\{X_1(t_0), X_2(t_0)\}$, where

$$\begin{aligned} W(X_1, X_2)(t_0) &= X_1(t_0)\dot{X}_2(t_0) - X_2(t_0)\dot{X}_1(t_0) \\ &= \begin{cases} \frac{-2}{|1-t_0^2|}, & t_0 \in (-1,1), \\ \frac{2}{|1-t_0^2|}, & t_0 \in (1,3). \end{cases} \end{aligned} \tag{8.9}$$

Proof. At each outcome $\omega \in \Omega$, the pair $\{X_1(t)(\omega), X_2(t)(\omega)\}$ is a fundamental set in the sample-path sense (because $X_1(t)$ and $X_2(t)$ are linearly independent $\mathrm{L}^\infty(\Omega)$ solutions), so we can compute the Wronskian pointwise on ω by using the deterministic Liouville's formula [41, Prop. 2.15]: if $t \in I$, then

$$W(X_1, X_2)(t)(\omega) = C_I(\omega) \mathrm{e}^{\int \frac{2t}{1-t^2}\,\mathrm{d}t} = \frac{C_I(\omega)}{|1-t^2|},$$

for a certain random variable $C_I(\omega)$ that depends on I. To obtain C_I at each $\omega \in \Omega$, notice that

$$\begin{aligned} C_I &= |1-t^2| W(X_1, X_2)(t) \\ &= |1-t^2| \left\{ X_1(t) \left(\dot{X}_1(t) \log|t-1| + \frac{X_1(t)}{t-1} + \sum_{n=1}^{\infty} n D_n (t-1)^{n-1} \right) - X_2(t)\dot{X}_1(t) \right\} \\ &\to \begin{cases} -2, & \text{if } t \to 1^-, \\ 2, & \text{if } t \to 1^+. \end{cases} \end{aligned}$$

This proves (8.9).
Note also that the random variables A_1 and A_2 defined by (8.8) belong to $\mathrm{L}^p(\Omega)$, because $X_1(t_0), \dot{X}_1(t_0), X_2(t_0), \dot{X}_2(t_0) \in \mathrm{L}^\infty(\Omega)$ and $Y_0, Y_1 \in \mathrm{L}^p(\Omega)$. Since $X_1(t)$ and $X_2(t)$ are L^∞ solutions on $(-1,1) \cup (1,3)$ by Proposition 8.1, and $A_1, A_2 \in \mathrm{L}^p(\Omega)$, from (8.7) we derive that $X(t)$ is an L^p solution to (8.1) on I.
To demonstrate the uniqueness, we use [160, Th. 5.1.2]. Rewrite (8.1) as a first-order linear differential equation

$$\dot{Z}(t) = B(t) Z(t),$$

where

$$Z(t) = \begin{pmatrix} X(t) \\ \dot{X}(t) \end{pmatrix}, \quad B(t) = \begin{pmatrix} 0 & 1 \\ \frac{A(A+1)}{1-t^2} & \frac{-2t}{1-t^2} \end{pmatrix}.$$

We say that $Z = (Z_1, Z_2)$ belongs to $\mathrm{L}_2^p(\Omega)$ if

$$\|Z\|_p \coloneqq \max\{\|Z_1\|_p, \|Z_2\|_p\} < \infty.$$

Consider the random matrix norm

$$|||B||| \coloneqq \max_i \sum_j \|b_{ij}\|_\infty.$$

If $Z, Z' \in \mathrm{L}_2^p(\Omega)$, then

$$\|B(t)Z - B(t)Z'\|_p \le |||B(t)||| \cdot \|Z - Z'\|_p,$$

where

$$\int_a^b |||B(t)||| \, \mathrm{d}t = \int_a^b \frac{\|A\|_\infty(\|A\|_\infty + 1) + 2|t|}{1 - t^2} \, \mathrm{d}t < \infty$$

for each $a < b$ that belong to I. Then the assumptions of [160, Th. 5.1.2] hold. □

To finish this section, we would like to comment that the hypothesis $A \in \mathrm{L}^\infty(\Omega)$ is not restrictive in practice, as any unbounded random variable can be truncated in a support as large as we want [125].

8.3 Approximation of the moments of the response: expectation and variance

Apart from determining the solution stochastic process $X(t)$ to the Legendre random differential equation (8.1), a main goal is also to construct reliable approximations regarding its statistical behavior. This latter information is mainly summarized by the mean and the variance functions. The mean or expectation function, $\mathbb{E}[X(t)]$, provides a measure of the average behavior of the process at each time instant t, while the variance, $\mathbb{V}[X(t)]$, quantifies the dispersion or variability of the process around the mean $\mathbb{E}[X(t)]$.

As the solution stochastic process $X(t)$ has been constructed via an infinite series, it is natural to approximate both its mean and its variance by considering truncations of that series to keep the computational burden affordable.

Now, we will show how under the conditions of Theorem 8.2, the moments of $X(t)$ can be approximated up to order p. We consider the truncation

$$X^N(t) = A_1^N X_1^N(t) + A_2^N X_2^N(t), \tag{8.10}$$

for $N \geq 1$, where

$$X_1^N(t) = \sum_{n=0}^{N} C_n(t-1)^n, \quad X_2^N(t) = X_1^N(t) \log|t-1| + \sum_{n=1}^{N} D_n(t-1)^n, \tag{8.11}$$

and

$$A_1^N = \frac{Y_0 \dot{X}_2^N(t_0) - Y_1 X_2^N(t_0)}{W(X_1, X_2)(t_0)}, \quad A_2^N = \frac{Y_1 X_1^N(t_0) - Y_0 \dot{X}_1^N(t_0)}{W(X_1, X_2)(t_0)}, \tag{8.12}$$

being $W(X_1, X_2)(t_0)$ the Wronskian computed in expression (8.9). As a consequence of Proposition 8.1, $X_1^N(t) \to X_1(t)$ and $X_2^N(t) \to X_2(t)$ in $\mathrm{L}^\infty(\Omega)$

as $N \to \infty$. Since $Y_0, Y_1 \in \mathrm{L}^p(\Omega)$, one has that $A_1^N \to A_1$ and $A_2^N \to A_2$ in $\mathrm{L}^p(\Omega)$ as $N \to \infty$. This implies that $X^N(t) \to X(t)$ in $\mathrm{L}^p(\Omega)$ as $N \to \infty$ too. In particular, the statistical moments up to order p of $X^N(t)$ tend to those of $X(t)$.

For $p \geq 2$ arbitrary but fixed, then we can approximate the average of $X(t)$, $\mathbb{E}[X(t)]$, and the variance of $X(t)$, $\mathbb{V}[X(t)]$, by using

$$\mathbb{E}[X(t)] = \lim_{N\to\infty} \mathbb{E}[X^N(t)], \quad \mathbb{V}[X(t)] = \lim_{N\to\infty} \mathbb{V}[X^N(t)],$$

see [160, Th. 4.2.1, Th. 4.3.1].

8.4 Numerical experiments

In this section we illustrate our theoretical findings by means of several numerical examples performed in the software Mathematica®. We will choose specific probability distributions for the input random variables A, Y_0 and Y_1 and then we will approximate the expectation and the variance of the response stochastic process $X(t)$ by using different orders of truncation N in expressions (8.10)–(8.12). The reliability of the obtained results will be shown by comparing them with the results provided by the following two techniques for uncertainty quantification:

- Monte Carlo simulation [70], which consists in obtaining a number m of realizations for the random input parameters, say

$$A^{(1)}, \ldots, A^{(m)},$$
$$Y_0^{(1)}, \ldots, Y_0^{(m)},$$
$$Y_1^{(1)}, \ldots, Y_1^{(m)},$$

  and then solving each one of the corresponding deterministic Legendre differential equations,

$$\begin{cases} (1-t^2)\ddot{X}^{(i)}(t) - 2t\dot{X}^{(i)}(t) + A^{(i)}(A^{(i)}+1)X^{(i)}(t) = 0, \\ X^{(i)}(t_0) = Y_0^{(i)}, \\ \dot{X}^{(i)}(t_0) = Y_1^{(i)}, \end{cases}$$

  which gives a realization $X^{(i)}(t)$ (a sample path) of $X(t)$, for $1 \leq i \leq m$. The expectation and variance of $X(t)$ can be approximated as follows:

$$\mathbb{E}[X(t)] \approx \mu_m(t) = \frac{1}{m}\sum_{i=1}^{m} X^{(i)}(t)$$

and

$$\mathbb{V}[X(t)] \approx \frac{1}{m-1}\sum_{i=1}^{m}\Big(X^{(i)}(t)-\mu_m(t)\Big)^2.$$

Monte Carlo simulation requires many realizations or simulations of $X(t)$ to get accurately its statistics (the error convergence rate is inversely proportional to the square root of the number m of realizations), therefore the computational cost of the Monte Carlo simulation is higher than our method based on random series.

- A variation of generalized Polynomial Chaos (gPC) expansions for continuous stochastic systems with dependent and jointly absolutely continuous random inputs, which is a method described in [49] and applied in [26]. We consider the canonical bases

$$\mathcal{C}_1^m = \{1, A, A^2, \ldots, A^m\},$$
$$\mathcal{C}_2^m = \{1, Y_0, Y_0^2, \ldots, Y_0^m\},$$
$$\mathcal{C}_3^m = \{1, Y_1, Y_1^2, \ldots, Y_1^m\},$$

and the vector of random input coefficients, $\zeta = (A, Y_0, Y_1)$. We consider a simple tensor product to construct a basis of monomials of degree less than or equal to m:

$$\Xi^p = \{\phi_0(\zeta), \phi_1(\zeta), \ldots, \phi_p(\zeta)\},$$

where

$$\phi_0 = 1, \quad p = \binom{m+3}{3}, \quad \phi_i(\zeta) = A^{i_1}Y_0^{i_2}Y_1^{i_3},$$

where $i_1 + i_2 + i_3 \leq m$ and $i \leftrightarrow (i_1, i_2, i_3)$ in a bijective manner.

We impose a solution to the random initial value problem (8.1) of the form

$$X^p(t) = \sum_{i=0}^{p}\tilde{X}_i^p(t)\phi_i(\zeta),$$

where $\tilde{X}_i^p(t)$, $0 \leq i \leq p$, are deterministic functions to be found. The deterministic differential equation satisfied by the coefficients is the following:

$$R\frac{\mathrm{d}^2}{\mathrm{d}t^2}\tilde{X}^p(t) - \frac{2t}{1-t^2}R\frac{\mathrm{d}}{\mathrm{d}t}\tilde{X}^p(t) + N\tilde{X}^p(t) = 0,$$

with initial conditions

$$R\tilde{X}^p(t_0) = v,$$

$$R\frac{\mathrm{d}}{\mathrm{d}t}\tilde{X}^p(t_0) = w,$$

where

$$\tilde{X}^p(t) = \left(\tilde{X}_0^p(t), \ldots, \tilde{X}_p^p(t)\right)^{\mathrm{T}},$$

$$v = \left(\mathbb{E}[Y_0\phi_0(\zeta)], \ldots, \mathbb{E}[Y_0\phi_p(\zeta)]\right)^{\mathrm{T}}$$

and

$$w = \left(\mathbb{E}[Y_1\phi_0(\zeta)], \ldots, \mathbb{E}[Y_1\phi_p(\zeta)]\right)^{\mathrm{T}},$$

and R and N are the following matrices of size $(p+1)\times(p+1)$:

$$R = \begin{pmatrix} \mathbb{E}[(\phi_0(\zeta))^2] & \cdots & \mathbb{E}[\phi_0(\zeta)\phi_p(\zeta)] \\ \vdots & \ddots & \vdots \\ \mathbb{E}[\phi_p(\zeta)\phi_0(\zeta)] & \cdots & \mathbb{E}[(\phi_p(\zeta))^2] \end{pmatrix}$$

and

$$N = \frac{1}{1-t^2}\begin{pmatrix} \mathbb{E}[A(A+1)(\phi_0(\zeta))^2] & \cdots & \mathbb{E}[A(A+1)\phi_0(\zeta)\phi_p(\zeta)] \\ \vdots & \ddots & \vdots \\ \mathbb{E}[A(A+1)\phi_p(\zeta)\phi_0(\zeta)] & \cdots & \mathbb{E}[A(A+1)(\phi_p(\zeta))^2] \end{pmatrix},$$

respectively.

This system of differential equations can be solved via standard numerical techniques, such as the Runge-Kutta algorithm. This method provides mean square approximations to $X(t)$ with spectral convergence rate in general, although numerical errors may arise for large orders m of bases, due to ill-conditioning of the matrix R for large p.

It can be shown that the expectation and the variance of solution stochastic process $X(t)$ can be approximated using the following finite sums:

$$\mathbb{E}[X(t)] \approx \mathbb{E}[X^p(t)] = \sum_{i=0}^{p} \tilde{X}_i^p(t)e_i,$$

$$\mathbb{V}[X(t)] \approx \mathbb{V}[X^p(t)] = \sum_{i,j=0}^{p} \tilde{X}_i^p(t)\tilde{X}_j^p(t)(R_{ij} - e_ie_j),$$

where $e_i = \mathbb{E}[\phi_i(\zeta)]$.

Now, we show two examples. We want to highlight that in the first example we assume that the input random data of the Legendre differential equation are assumed to be statistically dependent (with a joint probability distribution), while in the second example the corresponding random data are independent. In this manner we show that our theoretical results are able to consider both situations in practice.

Example 8.3 We assume the following joint probability distribution for the random input data:

$$(A, Y_0, Y_1) \sim \text{Dirichlet}(5, 1, 2, 3),$$

with initial conditions at $t_0 = 0$. Since A, Y_0 and Y_1 are bounded random variables, Theorem 8.2 ensures that $X(t)$ is an $\mathrm{L}^\infty(\Omega)$ solution to random initial value problem (8.1) on the interval $(-1, 1)$.

In Table 8.1 and Table 8.2, approximations of $\mathbb{E}[X(t)]$ and $\mathbb{V}[X(t)]$, respectively, are performed with order of truncations $N = 25$ and $N = 26$. The results are compared with Monte Carlo simulation and gPC expansions. Observe that there is stabilization of the results for $t \geq 0$, while for $t < 0$, specially for t near -1, larger truncation orders may be required. This is because the initial conditions are located at 0 and the power series are centered around 1, so better approximations are expected around those points. The results obtained agree with Monte Carlo simulation and gPC expansions. Notice that more simulations are required in the Monte Carlo method due to its slowness of convergence. On the other hand, gPC expansions converge quickly due to spectral convergence.

t	$\mathbb{E}[X^{25}(t)]$	$\mathbb{E}[X^{26}(t)]$	MC $500,000$	gPC $m = 4$	gPC $m = 5$
-0.9	-0.174008	-0.174833	-0.184978	-0.184818	-0.184818
-0.5	-0.0140125	-0.0140143	-0.0140897	-0.0140191	-0.0140191
0	0.0909091	0.0909091	0.090896	0.0909091	0.0909091
0.5	0.180172	0.180172	0.180222	0.180172	0.180172
0.9	0.281694	0.281694	0.281855	0.281694	0.281694

Table 8.1: Approximation of the expectation of the solution stochastic process. Example 8.3.

Example 8.4 In this example, we assume that A, Y_0 and Y_1 are independent random variables with the following probability distributions:

$$A \sim \text{Beta}(2, 1), \quad Y_0 \sim \text{Normal}(1, 1), \quad Y_1 \sim \text{Exponential}(2).$$

t	$\mathbb{V}[X^{25}(t)]$	$\mathbb{V}[X^{26}(t)]$	MC $500,000$	gPC $m=4$	gPC $m=5$
-0.9	0.0241268	0.0242201	0.0253679	0.0253799	0.0253799
-0.5	0.0107702	0.0107703	0.0107672	0.0107704	0.0107704
0	0.00688705	0.00688705	0.00689240	0.00688705	0.00688705
0.5	0.00844482	0.00844482	0.00845436	0.00844483	0.00844483
0.9	0.0262702	0.0262702	0.0262890	0.0262703	0.0262703

Table 8.2: Approximation of the variance of the solution stochastic process. Example 8.3.

The initial conditions are set at the time instant $t_0 = 1.2$. Since Y_0 and Y_1 have absolute moments of any order $1 \leq p < \infty$ and A is a bounded random variable, Theorem 8.2 tells us that $X(t)$ is an L^p solution to (8.1) for each $1 \leq p < \infty$ on $(1, 3)$.

Table 8.3 and Table 8.4 show approximations for the mean and variance of $X(t)$ at orders of truncation $N = 25$ and $N = 26$. The results obtained are compared with Monte Carlo simulation and gPC expansions. We observe that stabilization of the approximations occur for $t \leq 2.5$. For $t = 2.9$, a larger order of truncation N is needed, because the initial time $t_0 = 1.2$ and the center point 1 are far from $t = 2.9$. The approximations for the expectation and variance agree with Monte Carlo simulation and gPC expansions. The estimates performed by the gPC method are accurate due to spectral convergence, whereas more realizations for the Monte Carlo method are needed to achieve higher accuracy.

t	$\mathbb{E}[X^{25}(t)]$	$\mathbb{E}[X^{26}(t)]$	MC $500,000$	gPC $m=4$	gPC $m=5$
1.2	1	1	0.996418	1	1
1.5	1.14660	1.14660	1.14286	1.14660	1.14660
2	1.37833	1.37833	1.37402	1.37833	1.37833
2.5	1.59841	1.59841	1.59347	1.59841	1.59841
2.9	1.76891	1.76723	1.76260	1.76803	1.76803

Table 8.3: Approximation of the expectation of the solution stochastic process. Example 8.4.

t	$\mathbb{V}[X^{25}(t)]$	$\mathbb{V}[X^{26}(t)]$	MC 500,000	gPC $m=4$	gPC $m=5$
1.2	1	1	1.00313	1	1
1.5	1.12843	1.12843	1.13185	1.12843	1.12843
2	1.53899	1.53899	1.54307	1.53899	1.53899
2.5	2.06676	2.06676	2.07152	2.06676	2.06676
2.9	2.55896	2.55733	2.56339	2.55810	2.55810

Table 8.4: Approximation of the variance of the solution stochastic process. Example 8.4.

8.5 Conclusions

In this chapter, we have constructed a fundamental set $\{X_1(t), X_2(t)\}$ of $\mathrm{L}^p(\Omega)$ solutions ($1 \leq p \leq \infty$) to the Legendre differential equation with uncertainties on the domain $(-1,1) \cup (1,3)$ via random power series centered at the regular-singular point $t = 1$. It has been assumed that the initial conditions at the point $t_0 \in (-1,1) \cup (1,3)$, Y_0 and Y_1, belong to $\mathrm{L}^p(\Omega)$, and that the random input A is a bounded random variable (which is equivalent to A having absolute moments that grow at most exponentially). Under these hypotheses, a theorem on existence and uniqueness of $\mathrm{L}^p(\Omega)$ solution $X(t)$ to the random Legendre differential equation on $(-1,1)$ and $(1,3)$ has been proved. This result is an extension of Chapter 2, which constructed an $\mathrm{L}^p(\Omega)$ power series solution to the random Legendre differential equation on $(-1,1)$ around the regular point 0. In order to perform uncertainty quantification for $X(t)$, we have proposed a truncation method to approximate $X(t)$ by simpler processes $X^N(t)$ in the $\mathrm{L}^p(\Omega)$ sense, so that the moments of $X(t)$ up to order p can be approximated by those of $X^N(t)$. In particular, if $p \geq 2$, the expectation and variance of $X(t)$ can be approximated.

The numerical experiments have shown examples in which we have approximated both statistics of $X(t)$. These examples have been devised to consider two important situations from a practical standpoint, namely, when the input random data are assumed to be dependent and independent random variables. The results obtained have been compared with Monte Carlo simulation and gPC expansions, showing full agreement.

Finally, we would like to point out that our method could be extended to other important second-order random differential equations with a regular-singular point.

Acknowledgements

This work has been supported by the Spanish Ministerio de Economía y Competitividad grant MTM2017–89664–P. I acknowledge the doctorate scholarship granted by Programa de Ayudas de Investigación y Desarrollo (PAID), Universitat Politècnica de València.

The main results of this chapter have been published in [23].

Chapter 9

Density function of random differential equations via finite difference schemes: a theoretical analysis of a random diffusion-reaction Poisson-type problem

A computational approach to approximate the probability density function of random differential equations is based on transformation of random variables and finite difference schemes. The theoretical analysis of this computational method has not been performed in the extant literature. In this chapter, we deal with a particular random differential equation: a random diffusion-reaction Poisson-type problem of the form $-u''(x) + \alpha u(x) = \phi(x)$, $x \in [0,1]$, *with boundary conditions* $u(0) = A$, $u(1) = B$. *Here,* α, A *and* B *are random variables and* $\phi(x)$ *is a stochastic process. The term* $u(x)$ *is a stochastic process that solves the random problem in the sample-path sense. Via a finite difference scheme, we approximate* $u(x)$ *with a sequence of stochastic processes in both the almost sure and* L^p *senses. This allows us to find mild conditions under which the probability density function of* $u(x)$ *can be approximated. Illustrative examples are included.*

9.1 Introduction

Random differential equations are differential equations where the input coefficients and initial/boundary conditions are random variables/stochastic processes. The solution is a stochastic process. To completely understand the random behavior of the solution, one needs to find its joint finite-dimensional distributions, however, in general, this is an impracticable task. A more feasible target consists in finding, or at least approximating, its probability density function (first finite-dimensional distributions) [160, Ch. 3]. Some recent contributions dealing with the computation of the probability density function of the solution of random differential equations can be found in [57, 87]. A computational method to approximate the density function is based on finite difference schemes [62]. The theoretical analysis of this computational approach has not been done in the extant literature. In this chapter, we want to perform a comprehensive theoretical analysis for a particular random differential equation: a randomized diffusion-reaction Poisson-type problem [137, p. 433],

$$\begin{cases} -u''(x) + \alpha u(x) = \phi(x), & x \in [0,1], \\ u(0) = A,\ u(1) = B. \end{cases} \tag{9.1}$$

The term $-u''$ models diffusion, the term αu models reaction and ϕ represents an external source [137, p. 432]. We assume an underlying complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is the sample space, which consists of outcomes that will be generically denoted by ω; $\mathcal{F}$ is the σ-algebra of events; and $\mathbb{P}$ is the probability measure. The term $\alpha \geq 0$ and the boundary values A and B are random variables in our probability space. The source term $\phi(x)$ is a stochastic process. We will omit the evaluation at the outcome ω, however, when necessary, we will write $\alpha(\omega)$, $A(\omega)$, $B(\omega)$ and $\phi(x,\omega)$. The term $u(x)$ is a stochastic process that solves (9.1) in the sample-path sense. When evaluating at the outcome ω, we will write $u(x,\omega)$.

Notation 9.1 *Given a measure space $(S, \mathcal{F}, \mu)$, where $\mathcal{F}$ is the σ-algebra and μ is the measure, we will use the notation $\mathrm{L}^p(S)$ for the p-integrable measurable mappings in the Lebesgue sense: $f : S \to \mathbb{R}$ such that $\|f\|_{\mathrm{L}^p(S)} := (\int_S |f|^p \,\mathrm{d}\mu)^{1/p} < \infty$ for $1 \leq p < \infty$ and $\|f\|_{\mathrm{L}^\infty(S)} = \inf\{\sup\{|f(x)| : x \in S \backslash N\} : \mu(N) = 0\} < \infty$ for $p = \infty$. The shorten notation a.e. and a.s. will stand for "almost every" or "almost everywhere" and "almost surely", respectively.*

Given an interval $I \subseteq \mathbb{R}$, the notation $\mathfrak{C}^p(I)$, $p \in \mathbb{N} \cup \{\infty\}$, means p times continuously differentiable on I. When $p = 0$, it means continuous on I, and

we will write $\mathfrak{C}(I)$. For $0 < \beta \leq 1$, the notation $C^{\beta}(I)$ stands for the Hölder class: $f \in C^{\beta}(I)$ if there exists a constant $k > 0$ such that $|f(x) - f(y)| \leq k|x-y|^{\beta}$, for all $x, y \in I$. Do not confound $\mathfrak{C}^1(I)$ (continuously differentiable) with $C^1(I)$ (Lipschitz continuous). For $p \in \mathbb{N}$ and $0 < \beta \leq 1$, the notation $C^{p,\beta}(I)$ means being $\mathfrak{C}^p(I)$, with the p-th derivative being in $C^{\beta}(I)$.

Given a matrix $\mathcal{A}$, we will denote its p-norm as $\|\mathcal{A}\|_p$, $1 \leq p \leq \infty$. The j-th column of the matrix will be denoted by $\mathcal{A}(:, j)$, and its transpose will be written as $\mathcal{A}^T$. The identity matrix of size M will be denoted by I_M.

Finally, given an absolutely continuous random variable X, its density function will be denoted by f_X.

For the sake of completeness, below we give sufficient conditions on the external source ϕ in order to guarantee that the random problem (9.1) has a unique solution in different stochastic senses, commonly used in the extant literature.

Proposition 9.2 *The following holds:*

(i) If ϕ has sample paths in $\mathrm{L}^2([0,1])$, then there is a unique process u with sample paths in the Sobolev space $\mathrm{H}^2(0,1)$ [12, Ch. 8] that solve (9.1).

(ii) If ϕ has sample paths in $\mathfrak{C}([0,1])$, then there is a unique process u with sample paths in $\mathfrak{C}^2([0,1])$ that solve (9.1).

(iii) If ϕ has sample paths in $C^{\beta}([0,1])$, for some $0 < \beta \leq 1$, then there exists a unique process u with sample paths in $C^{2,\beta}([0,1])$ that solve (9.1).

Proof. The three statements are direct consequence of the deterministic theory for differential equations. Part (i) is a consequence of [137, Prop. 8.1] [12, Prop. 8.16]. Part (ii) is a consequence of part (i) and [12, Remark 6, p. 204]. Part (iii) is a consequence of [95, Th. 11.3.2].

□

The main goal of this chapter is to analyze when $u(x)$ is an absolutely continuous random variable, for each x, and then to compute its probability density function. For this purpose, we will use a finite difference scheme to approximate the solution stochastic process $u(x)$.

9.2 Random finite difference scheme

Divide $[0,1]$ into M equidistant interior points $x_1,\dots,x_M$: $0 = x_0 < x_1 < \dots < x_M < x_{M+1} = 1$, $x_i = x_i^M = i/(M+1)$. Denote $h = 1/(M+1)$. The numerical scheme, based on discretizations of the second derivative, is $-\frac{1}{h^2}u_{i+1}^M + (\frac{2}{h^2}+\alpha)u_i^M - \frac{1}{h^2}u_{i-1}^M = \phi(x_i)$, for $1 \le i \le M$, $u_0^M = A$ and $u_{M+1}^M = B$. One expects $u_i^M \approx u(x_i)$. In matrix form, $\mathcal{A}u^M = c$, where

$$u^M = \begin{pmatrix} u_1^M \\ \vdots \\ u_M^M \end{pmatrix}, \quad c = \begin{pmatrix} \phi(x_1) + A/h^2 \\ \phi(x_2) \\ \vdots \\ \phi(x_{M-1}) \\ \phi(x_M) + B/h^2 \end{pmatrix},$$

$$\mathcal{A} = \alpha I_M + \frac{1}{h^2}L, \quad L = \begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{pmatrix}.$$

Although not explicitly written, $\mathcal{A}$, L, c, x_i and h depend on M. Since $\alpha \ge 0$, then $\mathcal{A}$ is invertible. Thus, $u^M = \mathcal{A}^{-1}c$.

Lemma 9.3 *The matrix $\mathcal{A}$ satisfies $\|\mathcal{A}^{-1}\|_\infty \le 1/8$.*

Proof. The matrix $\mathcal{A}$ is an M-matrix, in the sense of [123, p. 10]: $\mathcal{A}$ has its offdiagonal entries nonpositive and, for $r = (r_i)_{i=1}^M$ with $r_i = 1/4 - (x_i - 1/2)^2 > 0$, one has $(\mathcal{A}r)_i = 2 + \alpha r_i \ge 2$, $1 \le i \le M$. By [123, Lemma 5.3], $\|\mathcal{A}^{-1}\|_\infty \le \|r\|_\infty / \min_i(\mathcal{A}r)_i \le \frac{1/4}{2} = 1/8$.

□

Proposition 9.4 *Suppose ϕ has sample paths in $C^\beta([0,1])$, for certain $0 < \beta \le 1$. Let $x_0 \in [0,1]$. Let $\{i_M\}_{M=1}^\infty$ be a sequence of indexes, $i_M \in \{1,\dots,M\}$, such that $\lim_{M\to\infty} i_M/(M+1) = x_0$. Then $\lim_{M\to\infty} u_{i_M}^M = u(x_0)$ a.s. Moreover, the following rate of convergence holds: $|u(x_{i_M},\omega) - u_{i_M}^M(\omega)| \le C(\omega)/8 \cdot h^\beta$, where $C(\omega)$ is the Hölder constant of $u''(\cdot,\omega)$ on $[0,1]$.*

Proof. By Proposition 9.2 (iii), the sample paths of u belong to $C^{2,\beta}([0,1])$: $|u''(x,\omega) - u''(y,\omega)| \le C(\omega)|x-y|^\beta$, for all $x, y \in [0,1]$. Using Taylor's expansions,

$$u(x+h) = u(x) + u'(x)h + u''(\xi_{x,h})h^2/2,\ u(x-h) = u(x) - u'(x)h + u''(\eta_{x,h})h^2/2,$$

where $\xi_{x,h} \in (x, x+h)$ and $\eta_{x,h} \in (x-h, x)$. The local error of the numerical scheme is given by

$$\begin{aligned} E_L(x,h) &= \frac{-u(x-h)+2u(x)-u(x+h)}{h^2} + \alpha u(x) - \phi(x) \\ &= -\frac{1}{2}(u''(\xi_{x,h}) + u''(\eta_{x,h})) + u''(x) \\ &= \frac{u''(x)-u''(\xi_{x,h})}{2} + \frac{u''(x)-u''(\eta_{x,h})}{2}. \end{aligned}$$

Using the triangular inequality and Hölder's condition,

$$|E_L(x,h)| \le \frac{C(\omega)}{2}|x-\xi_{x,h}|^\beta + \frac{C(\omega)}{2}|x-\eta_{x,h}|^\beta \le C(\omega)h^\beta.$$

Let $u^s = (u(x_i))_{i=1}^M$, $f^h = (E_L(x_i,h))_{i=1}^M$ and the error $e^h = u^s - u^M$. From $\mathcal{A}u^M = c$ and $\mathcal{A}u^s - c = f^h$, we derive that $\mathcal{A}e^h = f^h$. If we denote by $\|\cdot\|_\infty$ the infinity norm for matrices, $\|e^h\|_\infty \le \|\mathcal{A}^{-1}\|_\infty \|f^h\|_\infty \le \|\mathcal{A}^{-1}\|_\infty C(\omega)h^\beta$. By Lemma 9.3, $\|\mathcal{A}^{-1}\|_\infty \le 1/8$, therefore $\|e^h\|_\infty \le C(\omega)/8 \cdot h^\beta$. This implies $|u(x_{i_M},\omega) - u_{i_M}^M(\omega)| \le \frac{C(\omega)}{8}h^\beta$. By continuity of the sample paths of u,

$$\lim_{M\to\infty} u(x_{i_M},\omega) = \lim_{M\to\infty} u(i_M/(M+1),\omega) = u(x_0,\omega),$$

so we conclude that $\lim_{M\to\infty} u_{i_M}^M(\omega) = u(x_0,\omega)$, as wanted.

□

Proposition 9.5 *Suppose that ϕ has sample paths in $C^\beta([0,1])$, for certain $0 < \beta \le 1$. Let $x_0 \in [0,1]$. Let $\{i_M\}_{M=1}^\infty$ be a sequence of indexes, $i_M \in \{1,\dots,M\}$, such that $\lim_{M\to\infty} i_M/(M+1) = x_0$. Let $1 \le p < \infty$. If*

$$S := \max\{\|A\|_{\mathrm{L}^{p+\epsilon}(\Omega)}, \|B\|_{\mathrm{L}^{p+\epsilon}(\Omega)}, \sup_{x\in[0,1]} \|\phi(x)\|_{\mathrm{L}^{p+\epsilon}(\Omega)}\} < \infty$$

for some $\epsilon > 0$, then $u(x_0) \in \mathrm{L}^p(\Omega)$ and $\lim_{M\to\infty} u_{i_M}^M = u(x_0)$ in $\mathrm{L}^p(\Omega)$.

Proof. By Proposition 9.4, $\lim_{M\to\infty} u_{i_M}^M = u(x_0)$ a.s. Then, by [167, Th. 2.4], it suffices to check that

$$\sup_{M\ge 1} \|u_{i_M}^M\|_{\mathrm{L}^{p+\epsilon}(\Omega)} < \infty. \tag{9.2}$$

Write $\mathcal{A} = (\alpha + 2/h^2)I_M - H$, where

$$H = \frac{1}{h^2}\begin{pmatrix} 0 & 1 & & & \\ 1 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & 0 & 1 \\ & & & 1 & 0 \end{pmatrix}.$$

Let $T = 1/(\alpha+2/h^2)H$. Then $\mathcal{A} = (\alpha+2/h^2)(I_M - T)$. The eigenvalues of $\mathcal{A}$, μ_k, are well-known [155, p. 59]: $\mu_k = \alpha + 2/h^2 \cdot (1 - \cos(k\pi h))$, $k = 1, \ldots, M$. From this, the eigenvalues of T are easily computable:

$$\eta_k = \frac{\frac{2}{h^2}\cos(k\pi h)}{\alpha + \frac{2}{h^2}}, \quad k = 1, \ldots, M.$$

Since $|\cos(k\pi h)| < 1$, we deduce that

$$\|T\|_2 = \max\{|\eta_k| : k = 1, \ldots, M\} < \frac{2/h^2}{\alpha + 2/h^2} \leq 1.$$

The inequality $\|T\|_2 < 1$ implies that the matrix $I_M - T$ is invertible, with $(I_M - T)^{-1} = \sum_{k=0}^{\infty} T^k$. As a consequence,

$$\mathcal{A}^{-1} = \frac{1}{\alpha + \frac{2}{h^2}}(I_M - T)^{-1} = \frac{1}{\alpha + \frac{2}{h^2}} \sum_{k=0}^{\infty} T^k = \frac{1}{\alpha + \frac{2}{h^2}} \sum_{k=0}^{\infty} \frac{H^k}{\left(\alpha + \frac{2}{h^2}\right)^k}.$$

We derive that each entry $(\mathcal{A}^{-1})_{ij} = \sum_{k=0}^{\infty} (H^k)_{ij}/(\alpha+2/h^2)^{k+1}$ increases when α decreases, therefore $(\mathcal{A}^{-1})_{ij}$ takes its maximum value at $\alpha = 0$. On the other hand,

$$\begin{aligned} u^M_{i_M} = (\mathcal{A}^{-1}c)_{i_M} &= \sum_{j=1}^{M} (\mathcal{A}^{-1})_{i_M j} c_j \\ &= \underbrace{\frac{A}{h^2}(\mathcal{A}^{-1})_{i_M 1}}_{V_1} + \underbrace{\frac{B}{h^2}(\mathcal{A}^{-1})_{i_M M}}_{V_2} + \underbrace{\sum_{j=1}^{M} (\mathcal{A}^{-1})_{i_M j}\phi(x_j)}_{V_3}. \end{aligned}$$

Taking into account (9.2), we bound V_1, V_2 and V_3 in $\mathrm{L}^{p+\epsilon}(\Omega)$. First, we bound the third term:

$$\begin{aligned} \|V_3\|_{\mathrm{L}^{p+\epsilon}(\Omega)} &\leq \sum_{j=1}^{M} \|(\mathcal{A}^{-1})_{i_M j}\phi(x_j)\|_{\mathrm{L}^{p+\epsilon}(\Omega)} \leq \sum_{j=1}^{M} (\mathcal{A}^{-1})_{i_M j}|_{\alpha=0} \|\phi(x_j)\|_{\mathrm{L}^{p+\epsilon}(\Omega)} \\ &\leq S \sum_{j=1}^{M} (\mathcal{A}^{-1})_{i_M j}|_{\alpha=0} \leq S\|\mathcal{A}^{-1}|_{\alpha=0}\|_\infty \leq \frac{S}{8}, \end{aligned}$$

by Lemma 9.3. We bound $\|V_1\|_{\mathrm{L}^{p+\epsilon}(\Omega)}$, $\|V_2\|_{\mathrm{L}^{p+\epsilon}(\Omega)}$. Let $x^M = (x_1^M, \ldots, x_M^M)^T$, $x_i^M = i/(M+1)$. Let $\overline{x^M} = (x_M^M, \ldots, x_1^M)^T$ be the vector x^M reversed. Notice that $L\overline{x^M} = (1, 0, \ldots, 0)^T$ and $Lx^M = (0, \ldots, 0, 1)^T$, therefore $L^{-1}(:, 1) = \overline{x^M}$

and $L^{-1}(:,M) = x^M$. Since $\mathcal{A}|_{\alpha=0} = (1/h^2)L$, we derive that $\mathcal{A}^{-1}|_{\alpha=0}(:,1) = h^2\overline{x^M}$ and $\mathcal{A}^{-1}|_{\alpha=0}(:,M) = h^2 x^M$. Thus,

$$\begin{aligned}\|V_1\|_{\mathrm{L}^{p+\epsilon}(\Omega)} &= \frac{1}{h^2}\|A(\mathcal{A}^{-1})_{i_M 1}\|_{\mathrm{L}^{p+\epsilon}(\Omega)} \leq \frac{(\mathcal{A}^{-1})_{i_M 1}|_{\alpha=0}}{h^2}\|A\|_{\mathrm{L}^{p+\epsilon}(\Omega)} \\ &\leq \frac{h^2 x^M_{M+1-i_M}}{h^2} S = \left(1 - \frac{i_M}{M+1}\right) S \leq C < \infty, \end{aligned} \tag{9.3}$$

for some $C > 0$, since the sequence $\{i_M/(M+1)\}_{M=1}^{\infty}$ is bounded. Analogously, $\|V_2\|_{\mathrm{L}^{p+\epsilon}(\Omega)} < \infty$. This proves (9.2).

□

Proposition 9.6 *Suppose that ϕ has sample paths in $C^{\beta}([0,1])$, for certain $0 < \beta \leq 1$. Let $x_0 \in [0,1]$. Let $\{i_M\}_{M=1}^{\infty}$ be a sequence of indexes, $i_M \in \{1,\ldots,M\}$, such that $\lim_{M\to\infty} i_M/(M+1) = x_0$. Let $1 \leq p < \infty$. If $A, B \in \mathrm{L}^p(\Omega)$ and $\sup_{x\in[0,1]} |\phi(x)| \leq Y$ a.s. for some random variable $Y \in \mathrm{L}^p(\Omega)$, then $u(x_0) \in \mathrm{L}^p(\Omega)$ and $\lim_{M\to\infty} u^M_{i_M} = u(x_0)$ in $\mathrm{L}^p(\Omega)$.*

Proof. As in (9.3), $|V_1| \leq |A|(1 - i_M/(M+1)) \leq C_1|A|$, for some constant $C_1 > 0$. Similarly, $|V_2| \leq |B|(i_M/(M+1)) \leq C_2|B|$, for some constant $C_2 > 0$. On the other hand,

$$|V_3| \leq \sum_{j=1}^{M} (\mathcal{A}^{-1})_{i_M j}|\phi(x_j)| \leq Y \sum_{j=1}^{M} (\mathcal{A}^{-1})_{i_M j} \leq Y\|\mathcal{A}^{-1}\|_{\infty} \leq \frac{Y}{8},$$

by Lemma 9.3. Thus, $|u^M_{i_M}| \leq |V_1| + |V_2| + |V_3| \leq C_1|A| + C_2|B| + Y/8 \in \mathrm{L}^p(\Omega)$. By the Dominated Convergence Theorem [135, p. 321], $u(x_0) \in \mathrm{L}^p(\Omega)$ and $\lim_{M\to\infty} u^M_{i_M} = u(x_0)$ in $\mathrm{L}^p(\Omega)$, as wanted.

□

9.3 Probability density function of the solution stochastic process

Write $c = Ad + g$, where $g = (\phi(x_1),\ldots,\phi(x_{M-1}),\phi(x_M) + B/h^2)^T$ and $d = (1/h^2, 0,\ldots,0)^T$. Then $u_i^M = (\mathcal{A}^{-1}c)_i = (\mathcal{A}^{-1}d)_i A + (\mathcal{A}^{-1}g)_i$. Our next task is to compute the probability density function of u_i^M.

Lemma 9.7 *Let A be an absolutely continuous random variable, independent of the random vector (Z_1, Z_2), where $Z_1 \neq 0$ a.s. Then $Z_1 A + Z_2$ is absolutely continuous, with density function $f_{Z_1A+Z_2}(z) = \mathbb{E}[f_A((z - Z_2)/Z_1)/|Z_1|]$.*

Suppose that A is absolutely continuous, and that A and (α, B, ϕ) are independent (i.e., for all $0 \leq y_1, \ldots, y_m \leq 1$, $m \in \mathbb{N}$, A and $(\alpha, B, \phi(y_1), \ldots, \phi(y_m))$ are independent). By [123, Th. 5.2], $\mathcal{A}^{-1}$ has nonnegative entries, so $(\mathcal{A}^{-1}d)_i \geq 0$. In fact, $\mathcal{A}$ is an irreducible matrix, because its entries on the superdiagonal and on the subdiagonal are nonzero. By [7, Th. 2.7, p. 141], the entries of $\mathcal{A}^{-1}$ are positive, so $(\mathcal{A}^{-1}d)_i > 0$. By Lemma 9.7,

$$f_{u_i^M}(u) = \mathbb{E}\left[f_A\left(\frac{1}{(\mathcal{A}^{-1}d)_i} \left\{ u - (\mathcal{A}^{-1}g)_i \right\} \right) \frac{1}{(\mathcal{A}^{-1}d)_i} \right].$$

In practice, we can use an explicit expression for $f_{u_i^M}(u)$ that does not require the computation of $\mathcal{A}^{-1}$. The set of eigenvalues, μ_k, and eigenvectors, s^k, of $\mathcal{A}$ are known [155, p. 59]: $\mu_k = \alpha + 2/h^2 \cdot (1 - \cos(k\pi h))$, $s^k = (\sin(k\pi jh))_{j=1}^M$, $k = 1, \ldots, M$. Let $D = \operatorname{diag}(\mu_1, \ldots, \mu_M)$ and $P = [s^1 \ldots s^M]$. Since $s^1, \ldots, s^M$ are pairwise orthogonal and $\|s_i\|_2 = \sqrt{(M+1)/2}$, the matrix $R = \sqrt{2/(M+1)}\, P$ is orthogonal. We have the decomposition $\mathcal{A} = RDR^T$. Its inverse is given by $\mathcal{A}^{-1} = RD^{-1}R^T$. In the end, $f_{u_i^M}$ can be expressed as follows:

$$\begin{aligned} f_{u_i^M}(u) = \mathbb{E}\Bigg[f_A\Bigg(& \frac{h^2}{\frac{2}{M+1}\sum_{k=1}^M \frac{\sin(k\pi ih)\sin(k\pi h)}{\alpha + 2/h^2(1-\cos(k\pi h))}} \cdot \\ & \cdot \Bigg\{ u - \frac{2}{M+1} \sum_{j=1}^M \left(\sum_{k=1}^M \frac{\sin(k\pi ih)\sin(k\pi jh)}{\alpha + 2/h^2(1-\cos(k\pi h))} \right) \phi(x_j) \\ & - \frac{2}{M+1} \frac{B}{h^2} \sum_{k=1}^M \frac{\sin(k\pi ih)\sin(k\pi Mh)}{\alpha + 2/h^2(1-\cos(k\pi h))} \Bigg\} \Bigg) \frac{h^2}{\frac{2}{M+1}\sum_{k=1}^M \frac{\sin(k\pi ih)\sin(k\pi h)}{\alpha + 2/h^2(1-\cos(k\pi h))}} \Bigg]. \end{aligned} \tag{9.4}$$

Theorem 9.8 *Suppose that ϕ has sample paths in $C^\beta([0,1])$ (certain $0 < \beta \leq 1$), A is absolutely continuous, A and (B, α, ϕ) are independent, f_A is continuous and bounded on $\mathbb{R}$ and $\|\alpha\|_{\mathrm{L}^\infty(\Omega)} < \infty$. Let $x_0 \in (0,1)$. Let $\{i_M\}_{M=1}^\infty$ be a sequence of indexes, $i_M \in \{1, \ldots, M\}$, such that $\lim_{M\to\infty} i_M/(M+1) = x_0$. Then the sequence $\{f_{u_{i_M}^M}(u)\}_{M=1}^\infty$, defined in (9.4), converges to a density $f_{u(x_0)}(u)$ of $u(x_0)$, for all $u \in \mathbb{R}$.*

Proof. By Proposition 9.2, let v and w be two stochastic processes with sample paths in $C^{2,\beta}([0,1])$ that solve

$$\begin{cases} -v''(x) + \alpha v(x) = 0, \ x \in [0,1], \\ v(0) = 1, \ v(1) = 0, \end{cases} \qquad \begin{cases} -w''(x) + \alpha w(x) = \phi(x), \ x \in [0,1], \\ w(0) = 0, \ w(1) = B. \end{cases}$$

By Proposition 9.4, $(\mathcal{A}^{-1}d)_{i_M} \to v(x_0)$ and $(\mathcal{A}^{-1}g)_{i_M} \to w(x_0)$ a.s. as $M \to \infty$. Since f_A is continuous,

$$\lim_{M\to\infty} f_A\left(\frac{1}{(\mathcal{A}^{-1}d)_{i_M}}\left\{u-(\mathcal{A}^{-1}g)_{i_M}\right\}\right)\frac{1}{(\mathcal{A}^{-1}d)_{i_M}} = f_A\left(\frac{1}{v(x_0)}\left\{u-w(x_0)\right\}\right)\frac{1}{v(x_0)} \text{ a.s.} \tag{9.5}$$

Note that it makes sense to divide by $v(x_0)$, since by the Strong maximum principle for elliptic PDEs [67, Th. 4, p. 333], $0 < v(x) < 1$ for all $x \in (0,1)$, a.s.

Claim: there exists M_0 (independent of ω) such that, for all $M \geq M_0$, $(\mathcal{A}^{-1}d)_{i_M} \geq v(x_0)/2$ a.s.

We prove the claim. The random boundary value problem satisfied by v can be solved analytically. Indeed, fixed $\omega \in \Omega$, we distinguish two cases according to $\alpha(\omega) > 0$ or $\alpha(\omega) = 0$:

$\alpha(\omega) > 0$: The solution is $v(x) = \sinh(\sqrt{\alpha}(1-x))/\sinh(\sqrt{\alpha})$. Then $v''(x) = \alpha v(x) = \alpha\sinh(\sqrt{\alpha}(1-x))/\sinh(\sqrt{\alpha})$. By the Mean Value Theorem, $|v''(x)-v''(y)| = \frac{\alpha^2}{\sinh(\sqrt{\alpha})}\cosh(\xi_{x,y,\omega})|x-y|$, where $\xi_{x,y,\omega} \leq \max\{\sqrt{\alpha}(1-x), \sqrt{\alpha}(1-y)\} \leq \sqrt{\alpha}$. Hence,

$$|v''(x)-v''(y)| \leq \frac{\alpha^2\cosh(\sqrt{\alpha})}{\sinh(\sqrt{\alpha})}|x-y| \leq \frac{\|\alpha\|^2_{\mathrm{L}^\infty(\Omega)}\cosh(\sqrt{\|\alpha\|_{\mathrm{L}^\infty(\Omega)}})}{\sinh(\sqrt{\|\alpha\|_{\mathrm{L}^\infty(\Omega)}})}|x-y|.$$

$\alpha(\omega) = 0$: The solution is $v(x) = 1-x$, so $|v''(x)-v''(y)| = 0\cdot|x-y|$.

Let

$$K = \begin{cases} \frac{\|\alpha\|^2_{\mathrm{L}^\infty(\Omega)}\cosh(\sqrt{\|\alpha\|_{\mathrm{L}^\infty(\Omega)}})}{\sinh(\sqrt{\|\alpha\|_{\mathrm{L}^\infty(\Omega)}})}, & \|\alpha\|_{\mathrm{L}^\infty(\Omega)} > 0, \\ 0, & \|\alpha\|_{\mathrm{L}^\infty(\Omega)} = 0. \end{cases}$$

Thus, $|v''(x)-v''(y)| \leq K|x-y|$. Therefore, the Hölder constant $C(\omega)$ of $v''(\cdot,\omega)$ can be taken independently of ω: $C(\omega) = K$. By Proposition 9.4, $|v(x_{i_M},\omega) - v^M_{i_M}(\omega)| \leq K/8\cdot h$ a.s., where $v^M_{i_M} = (\mathcal{A}^{-1}d)_{i_M}$. On the other hand, using the Mean Value Theorem to estimate $|v(x_0)-v(x_{i_M})|$ as we did before, we obtain $|v(x_0,\omega)-v(x_{i_M},\omega)| \leq L|x_0-x_{i_M}|$ a.s., where

$$L = \begin{cases} \frac{\|\alpha\|_{\mathrm{L}^\infty(\Omega)}\cosh(\sqrt{\|\alpha\|_{\mathrm{L}^\infty(\Omega)}})}{\sinh(\sqrt{\|\alpha\|_{\mathrm{L}^\infty(\Omega)}})}, & \|\alpha\|_{\mathrm{L}^\infty(\Omega)} > 0, \\ 1, & \|\alpha\|_{\mathrm{L}^\infty(\Omega)} = 0. \end{cases}$$

By the triangular inequality,

$$|v(x_0,\omega) - v^M_{i_M}(\omega)| \le |v(x_0,\omega) - v(x_{i_M},\omega)| + |v(x_{i_M},\omega) - v^M_{i_M}(\omega)| \\ \le K/8 \cdot h + L|x_0 - x_{i_M}| \text{ a.s.}$$

Now, if $\alpha(\omega) > 0$, we know that $v(x_0) = \sinh(\sqrt{\alpha}(1-x_0))/\sinh(\sqrt{\alpha})$. As a function of α, it has a lower bound $m > 0$. Then, $v(x_0,\omega) \ge m$ a.s. To conclude, take M_0 such that, for all $M \ge M_0$, $K/8 \cdot h + L|x_0 - x_{i_M}| \le m/2$. This implies that $|v(x_0,\omega) - v^M_{i_M}(\omega)| \le m/2 \le v(x_0,\omega)/2$, therefore $v^M_{i_M}(\omega) \ge v(x_0,\omega)/2$ a.s. This concludes the proof of the claim.
Hence, for $M \ge M_0$,

$$f_A\left(\frac{1}{(\mathcal{A}^{-1}d)_{i_M}}\left\{u - (\mathcal{A}^{-1}g)_{i_M}\right\}\right)\frac{1}{(\mathcal{A}^{-1}d)_{i_M}} \le \|f_A\|_{\mathrm{L}^\infty(\mathbb{R})}\frac{2}{v(x_0)} \le \|f_A\|_{\mathrm{L}^\infty(\mathbb{R})}\frac{2}{m}.$$

Since $\|f_A\|_{\mathrm{L}^\infty(\mathbb{R})}2/m$ is constant, it belongs to $\mathrm{L}^1(\Omega)$. By the Dominated Convergence Theorem [135, p. 321],

$$\lim_{M\to\infty} f_{u^M_{i_M}}(u) = \mathbb{E}\left[f_A\left(\frac{1}{v(x_0)}\left\{u - w(x_0)\right\}\right)\frac{1}{v(x_0)}\right] =: \bar{f}(u).$$

Finally, we prove that $\bar{f}$ is a density of $u(x_0)$. Let G be a random variable with density function given by $\bar{f}$. By Scheffé's Lemma [170, p. 55], $u^M_{i_M} \to G$ in law as $M \to \infty$. On the other hand, by Proposition 9.4, $u^M_{i_M} \to u(x_0)$ a.s., so $u^M_{i_M} \to u(x_0)$ in law, as $M \to \infty$. Then $u(x_0)$ and G are equal in distribution, so $f_{u(x_0)}(u) = \bar{f}(u)$, as wanted.

□

The continuity of f_A on $\mathbb{R}$ is satisfied, for instance, by the density function of the distributions Normal(μ,σ^2), $\mu \in \mathbb{R}$ and $\sigma^2 > 0$; Beta(a,b), $a > 1$ and $b > 1$; Gamma(a,b), $a > 1$ and $b > 0$; etc. However, it would be desirable to require only a.e. continuity for f_A, since the class of applicable density functions would be larger: Beta(a,b), $a \ge 1$ and $b \ge 1$; Uniform(a,b), $a < b$; Gamma(a,b), $a \ge 1$ and $b > 0$ (in particular, Exponential(b)); truncated normal distribution; etc. This is the purpose of the following theorem.

Theorem 9.9 *Suppose that ϕ has sample paths in $C^\beta([0,1])$ (certain $0 < \beta \le 1$), A, B and α are absolutely continuous, A, B and (α,ϕ) are independent, $\|\alpha\|_{\mathrm{L}^\infty(\Omega)} < \infty$, f_A is a.e. continuous and essentially bounded on $\mathbb{R}$. Let $x_0 \in (0,1)$. Let $\{i_M\}_{M=1}^\infty$ be a sequence of indexes, $i_M \in \{1,\dots,M\}$, such that $\lim_{M\to\infty} i_M/(M+1) = x_0$. Then the sequence $\{f_{u^M_{i_M}}(u)\}_{M=1}^\infty$, defined in (9.4), converges to a density $f_{u(x_0)}(u)$ of $u(x_0)$, for all $u \in \mathbb{R}$.*

Proof. The proof is analogous to Theorem 9.8. Since $\alpha(\omega) > 0$ a.s. (because $\mathbb{P}(\alpha = 0) = 0$), we have $v(x_0) = \sinh(\sqrt{\alpha}(1 - x_0))/\sinh(\sqrt{\alpha})$ a.s. By the Random Variable Transformation (RVT) technique [47, Th. 1] (briefly, it is a method that consists in computing the probability density function of a transformation of an absolutely continuous random variable/vector), $v(x_0)$ is absolutely continuous. On the other hand, w can also be explicitly found using the theory of linear differential equations, and one obtains that $w(x_0)$ can be written as $Z_1 B + Z_2$, where B and (Z_1, Z_2) are independent and $Z_1 \neq 0$ a.s. By Lemma 9.7, $w(x_0)$ is absolutely continuous. By the RVT technique, $1/v(x_0) \cdot \{u - w(x_0)\}$ is absolutely continuous. Then, the probability that $1/v(x_0) \cdot \{u - w(x_0)\}$ belongs to the discontinuity set of f_A is 0. By the Continuous Mapping Theorem [165, p. 7, Th. 2.3], (9.5) holds. The rest of the proof is as in Theorem 9.8.

□

In the following proposition, we study whether the finite difference scheme preserves the pointwise convergence of the derivatives.

Proposition 9.10 *Assume the conditions of Theorem 9.8. If the n-th derivative $f_A^{(n)}$ exists on $\mathbb{R}$ and $f_A^{(j)}$ is bounded on $\mathbb{R}$ for each $1 \leq j \leq n$, then $f_{u_{i_M}^M}$ and $f_{u(x_0)}$ have bounded n-th derivatives on $\mathbb{R}$. Moreover, if $f_A^{(n)}$ is continuous on $\mathbb{R}$, then the sequence $\{f_{u_{i_M}^M}^{(n)}(u)\}_{M=1}^{\infty}$ converges to $f_{u(x_0)}^{(n)}(u)$, for all $u \in \mathbb{R}$.*

Proof. Both $f_{u_{i_M}^M}$ and $f_{u(x_0)}$ possess n-th derivatives on $\mathbb{R}$ because of the differentiability of f_A and the Dominated Convergence Theorem. Indeed, fix $u \in \mathbb{R}$. We have

$$\frac{f_A\left(\frac{u+h-w(x_0)}{v(x_0)}\right)\frac{1}{v(x_0)} - f_A\left(\frac{u-w(x_0)}{v(x_0)}\right)\frac{1}{v(x_0)}}{h} \xrightarrow{h\to 0} f_A'\left(\frac{u - w(x_0)}{v(x_0)}\right)\frac{1}{v(x_0)^2}, \quad (9.6)$$

by definition of derivative. Now, by the deterministic Mean Value Theorem,

$$\begin{aligned}\left|\frac{f_A\left(\frac{u+h-w(x_0)}{v(x_0)}\right)\frac{1}{v(x_0)} - f_A\left(\frac{u-w(x_0)}{v(x_0)}\right)\frac{1}{v(x_0)}}{h}\right| &= \left|f_A'\left(\frac{u + \xi_h - w(x_0)}{v(x_0)}\right)\right|\frac{1}{v(x_0)^2} \\ &\leq \|f_A'\|_{\mathrm{L}^\infty(\mathbb{R})}\frac{1}{m^2}, \quad (9.7)\end{aligned}$$

where ξ_h depends on ω and $|\xi_h| < |h|$. Notice that the inequality $v(x_0) \geq m$ a.s. from the proof of Theorem 9.8 has been utilized. The Dominated Convergence

Theorem thus applies to ensure the existence of

$$f'_{u(x_0)}(u) = \mathbb{E}\left[f'_A\left(\frac{1}{v(x_0)}\left\{u - w(x_0)\right\}\right)\frac{1}{v(x_0)^2}\right]. \tag{9.8}$$

Analogously one justifies that

$$f'_{u^M_{i_M}}(u) = \mathbb{E}\left[f'_A\left(\frac{1}{(\mathcal{A}^{-1}d)_{i_M}}\left\{u - (\mathcal{A}^{-1}g)_{i_M}\right\}\right)\frac{1}{\left((\mathcal{A}^{-1}d)_{i_M}\right)^2}\right] \tag{9.9}$$

exists, by using the fact that $(\mathcal{A}^{-1}d)_{i_M} \geq v(x_0)/2 \geq m/2$ a.s. For higher derivatives, the procedure works analogously and one proves

$$f^{(n)}_{u(x_0)}(u) = \mathbb{E}\left[f^{(n)}_A\left(\frac{1}{v(x_0)}\left\{u - w(x_0)\right\}\right)\frac{1}{v(x_0)^n}\right], \tag{9.10}$$

$$f^{(n)}_{u^M_{i_M}}(u) = \mathbb{E}\left[f^{(n)}_A\left(\frac{1}{(\mathcal{A}^{-1}d)_{i_M}}\left\{u - (\mathcal{A}^{-1}g)_{i_M}\right\}\right)\frac{1}{\left((\mathcal{A}^{-1}d)_{i_M}\right)^n}\right]. \tag{9.11}$$

When $f^{(n)}_A$ is continuous on $\mathbb{R}$, the proof of Theorem 9.8 works with $f^{(n)}_A$ instead of f_A, so that one concludes that $\{f^{(n)}_{u^M_{i_M}}(u)\}_{M=1}^{\infty}$ converges to $f^{(n)}_{u(x_0)}(u)$, for all $u \in \mathbb{R}$.

□

Remark 9.11 The same analysis can be performed if B is absolutely continuous and B and (A, α, ϕ) are independent. In such a case,

$$f_{u^M_i}(u) = \mathbb{E}\Bigg[f_B\Bigg(\frac{h^2}{\frac{2}{M+1}\sum_{k=1}^{M}\frac{\sin(k\pi ih)\sin(k\pi Mh)}{\alpha+2/h^2(1-\cos(k\pi h))}}\cdot$$
$$\cdot\Bigg\{u - \frac{2}{M+1}\sum_{j=1}^{M}\left(\sum_{k=1}^{M}\frac{\sin(k\pi ih)\sin(k\pi jh)}{\alpha+2/h^2(1-\cos(k\pi h))}\right)\phi(x_j)$$
$$-\frac{2}{M+1}\frac{A}{h^2}\sum_{k=1}^{M}\frac{\sin(k\pi ih)\sin(k\pi h)}{\alpha+2/h^2(1-\cos(k\pi h))}\Bigg\}\Bigg)\frac{h^2}{\frac{2}{M+1}\sum_{k=1}^{M}\frac{\sin(k\pi ih)\sin(k\pi Mh)}{\alpha+2/h^2(1-\cos(k\pi h))}}\Bigg].$$

One could think of performing the same analysis by isolating $\phi(x_1)$, instead of A or B. In such a case, one would assume that $\phi(x_1)$ is absolutely continuous, and that $(\alpha, A, B, \phi(x_2), \ldots, \phi(x_M))$ and $\phi(x_1)$ are independent. To achieve this independence, one may require $\phi(y_1), \ldots \phi(y_m)$ to be independent, for every $y_1, \ldots, y_m \in [0,1]$, $m \geq 1$. A process ϕ of this type exists by Kolmogorov's Extension Theorem [9, Th. 36.2, p. 486]. However, by [96, Example 1.2.5,

p. 10], this process ϕ is not jointly measurable on $[0,1]\times\Omega$. This implies that its sample paths cannot be right-continuous nor left-continuous, so ϕ does not have enough regularity to apply our results.

Remark 9.12 The theoretical expression of $f_{u_i^M}(u)$ is

$$
\begin{aligned}
f_{u_i^M}(u) = \int_{\mathbb{R}^{M+2}} f_A\Bigg(& \frac{h^2}{\frac{2}{M+1}\sum_{k=1}^{M}\frac{\sin(k\pi ih)\sin(k\pi h)}{\alpha+2/h^2(1-\cos(k\pi h))}}\cdot \\
& \cdot\Bigg\{ u - \frac{2}{M+1}\sum_{j=1}^{M}\left(\sum_{k=1}^{M}\frac{\sin(k\pi ih)\sin(k\pi jh)}{\alpha+2/h^2(1-\cos(k\pi h))}\right)\phi_j \\
& -\frac{2}{M+1}\frac{b}{h^2}\sum_{k=1}^{M}\frac{\sin(k\pi ih)\sin(k\pi Mh)}{\alpha+2/h^2(1-\cos(k\pi h))}\Bigg\}\Bigg)\frac{h^2}{\frac{2}{M+1}\sum_{k=1}^{M}\frac{\sin(k\pi ih)\sin(k\pi h)}{\alpha+2/h^2(1-\cos(k\pi h))}} \\
& \cdot\mathbb{P}_{(\alpha,B,\phi(x_1),\ldots,\phi(x_M))}(\mathrm{d}\alpha,\mathrm{d}b,\mathrm{d}\phi_1,\ldots,\mathrm{d}\phi_M).
\end{aligned}
$$

However, in practice, we use Monte Carlo simulation to compute the expectation (9.4), by sampling from α, B and ϕ.

9.4 Examples and Conclusions

Example 9.13 Consider (9.1) with $A\sim\text{Gamma}(2,1)$, $B\sim\text{Poisson}(3)$, $\alpha\sim\text{Uniform}(1,2)$ and $\phi(x)=\arctan(\mathrm{e}^{\cos(Dx)}+1)$, where $D\sim\text{Binomial}(20,0.2)$. The random variables are assumed to be independent. Notice that ϕ has sample paths in $\mathfrak{C}^\infty([0,1])$, therefore in $C^\beta([0,1])$. Then there is a solution process u in $C^{2,\beta}([0,1])$ (in fact, in $\mathfrak{C}^\infty([0,1])$). On the other hand, f_A is continuous and bounded on $\mathbb{R}$. Then Theorem 9.8 allows us to approximate the density function of $u(x)$, $x\in(0,1)$. Also, since A and B have moments of all orders and $|\phi(x)|\leq\pi/2$ a.s., both Proposition 9.5 and Proposition 9.6 ensure that $u(x)$ has moments of all orders and, moreover, they can be approximated. In particular, the expectation and variance of $u(x)$, $\mathbb{E}[u(x)]$ and $\mathbb{V}[u(x)]$, can be approximated.

We will do so for $x=0.5$. Let $i_M=(M+1)/2$, for M odd. Then $\{f_{u_{i_M}^M}(u)\}_{M\text{odd}}$, defined in (9.4), tends to $f_{u(0.5)}(u)$, $u\in\mathbb{R}$. In Figure 9.1, we show the graph of $f_{u_{i_M}^M}(u)$ for $M=9,11,13$. In Table 9.1, we compute

$$
\mathbb{E}[u_{i_M}^M]=\int_{\mathbb{R}} u f_{u_{i_M}^M}(u)\,\mathrm{d}u,\quad \mathbb{V}[u_{i_M}^M]=\int_{\mathbb{R}} u^2 f_{u_{i_M}^M}(u)\,\mathrm{d}u-\left(\mathbb{E}[u_{i_M}^M]\right)^2,
$$

for $M = 9, 11, 13$. We observe that there is convergence, which agrees with our theoretical findings, and moreover it is rapid. The expectation in (9.4) has been computed via Monte Carlo simulation, as explained in Remark 9.12, with 100,000 samples of the involved random variables for each M.

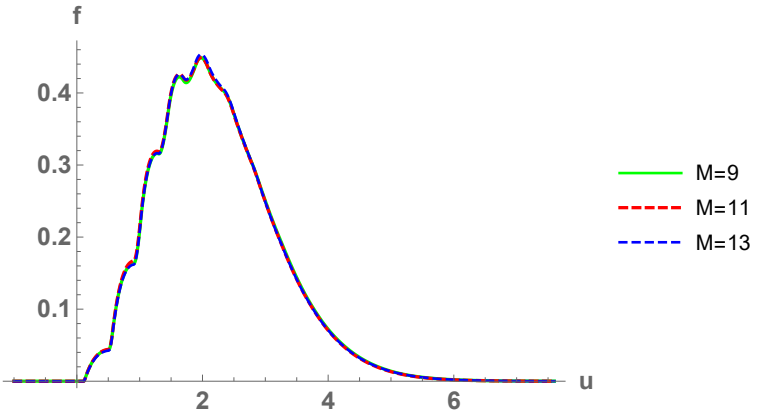

Figure 9.1: Graph of $f_{u_{i_M}^M}(u)$ for $M = 9$ (green), $M = 11$ (thick dashed red) and $M = 13$ (tiny dashed blue). Example 9.13.

M	9	11	13
$\mathbb{E}[u_{i_M}^M]$	2.22	2.21	2.21
$\mathbb{V}[u_{i_M}^M]$	0.89	0.88	0.88

Table 9.1: Expectation and variance of $u_{i_M}^M$, for $M = 9, 11, 13$. Example 9.13.

Example 9.14 We deal with (9.1) having as inputs $A \sim \text{Uniform}(-1, 1)$, $B \sim \text{Gamma}(4, 1)$, $\alpha \sim \text{Uniform}(1, 2)$ and $\phi(x)$ a standard Brownian motion on $[0, 1]$ ($\phi(x) \sim \text{Normal}(0, x)$) [119, Ch. 5]. The random variables/process are assumed to be independent. Brownian motion has $C^\beta([0, 1])$ sample paths, for $0 < \beta < 1/2$. By Proposition 9.2, there exists a solution process u in $C^{2,\beta}([0, 1])$. On the other hand, f_A is a.e. continuous (two points of discontinuity, -1 and 1) and bounded on $\mathbb{R}$, and A, B and α are absolutely continuous. Then Theorem 9.9 allows us to approximate the density function of $u(x)$, $x \in (0, 1)$. In addition, since A and B have moments of all orders and $\sup_{x\in[0,1]} \|\phi(x)\|_{\mathrm{L}^p(\Omega)} = \sup_{x\in[0,1]} (\frac{1}{\sqrt{\pi}}(2x)^{p/2}\Gamma(\frac{p+1}{2}))^{1/p} \leq (\frac{1}{\sqrt{\pi}}2^{p/2}\Gamma(\frac{p+1}{2}))^{1/p} < \infty$ for each $p \geq 1$, Proposition 9.5 entails that $u(x)$ has moments of all orders

and, moreover, they can be approximated. In particular, the expectation and variance of $u(x)$, $\mathbb{E}[u(x)]$ and $\mathbb{V}[u(x)]$, can be approximated.

We work at the point $x = 0.5$ again. Let $i_M = (M+1)/2$, for M odd. Then $\{f_{u_{i_M}^M}(u)\}_{M \text{ odd}}$, defined in (9.4), tends to $f_{u(0.5)}(u)$, $u \in \mathbb{R}$. In Figure 9.2, we depict the graph of $f_{u_{i_M}^M}(u)$ for $M = 9, 11, 13$. In Table 9.2, we calculate the expectation and variance for $M = 9, 11, 13$. We observe convergence, which agrees with our theoretical findings, and furthermore this convergence is fast. The expectation in (9.4) has been determined via Monte Carlo simulation, see Remark 9.12, with $100,000$ realizations of the involved random variables for each M. To sample from a Brownian motion, we use its Karhunen-Loève expansion on $[0, 1]$ with a sufficiently large order of truncation.

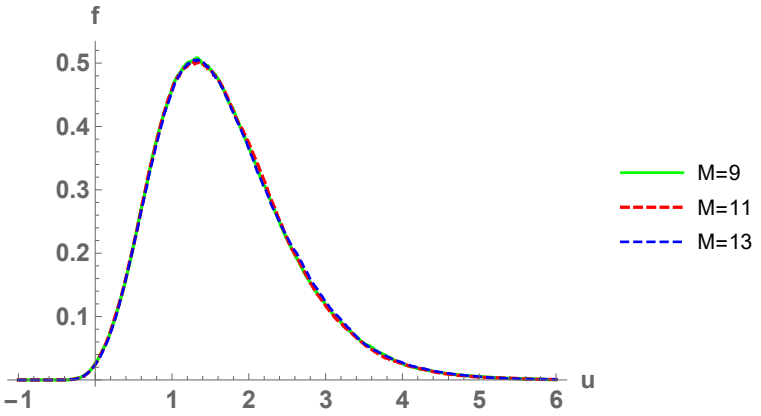

Figure 9.2: Graph of $f_{u_{i_M}^M}(u)$ for $M = 9$ (green), $M = 11$ (thick dashed red) and $M = 13$ (tiny dashed blue). Example 9.14.

M	9	11	13
$\mathbb{E}[u_{i_M}^M]$	1.68	1.68	1.68
$\mathbb{V}[u_{i_M}^M]$	0.77	0.77	0.77

Table 9.2: Expectation and variance of $u_{i_M}^M$, for $M = 9, 11, 13$. Example 9.14.

Finally, we want to point out that this study seeks to contribute to the field of random differential equations, where a main goal is to compute the mean and variance of the solution stochastic process. In this chapter we have shown

a novel method to go beyond the computation of these statistical moments. Indeed, we have rigorously addressed the computation of the probability density function of an important random diffusion-reaction problem with random boundary conditions, by using a finite difference numerical scheme. The proposed approach can be very useful to deal with other significant random differential equations.

Research on the rate of convergence of the approximating density functions to the target density function could be conducted in the future. An issue that should be resolved in such a case is the fact that our reasoning is entirely based on existing results of convergence in Probability and Analysis (Dominated Convergence Theorem, Continuous Mapping Theorem, [20, Th. 2.4], etc.), which, at least to our knowledge, do not usually provide rates of convergence. Thus, in order to obtain optimal or at least sub-optimal bounds, we should proceed with step-by-step inequalities. We believe that this might be achievable by assuming f_A to be Lipschitz continuous on $\mathbb{R}$ and by finding the constants involved in the proof of Theorem 9.8. See our recent contribution [19], in which some theoretical rates of convergence for the approximating density functions were found in the setting of a random parabolic partial differential equation. These ideas raise new research lines for the future.

Acknowledgements

This work has been supported by the Spanish Ministerio de Economía y Competitividad grant MTM2017–89664–P. I acknowledge the doctorate scholarship granted by Programa de Ayudas de Investigación y Desarrollo (PAID), Universitat Politècnica de València.

The main results of this chapter have been published in [15]. This chapter is a national collaboration with the researcher José Antonio Díaz Navas (Departamento de Óptica, Facultad de Ciencias, Universidad de Granada, Granada, Spain).

Chapter 10

L^p-calculus approach to the random autonomous linear differential equation with discrete delay

In this chapter, we provide a full probabilistic study of the random autonomous linear differential equation with discrete delay $\tau > 0$*:* $x'(t) = ax(t)+bx(t-\tau)$*,* $t \geq 0$*, with initial condition* $x(t) = g(t)$*,* $-\tau \leq t \leq 0$*. The coefficients* a *and* b *are assumed to be random variables, while the initial condition* $g(t)$ *is taken as a stochastic process. By using* L^p*-calculus, we prove that, under certain conditions, the deterministic solution constructed with the method of steps that involves the delayed exponential function is an* L^p*-solution too. An analysis of* L^p*-convergence when the delay* τ *tends to* 0 *is also performed in detail.*

10.1 Introduction

Delay differential equations can be viewed as generalizations of classical differential equations. The study of delay differential equations requires a distinctive treatment with respect to their classical counterpart [156]. This fact can be checked starting from introducing a delay in the basic linear ordinary differential equation that leads to richer qualitative and quantitative behaviors [60]. Regarding applications, the delays or lags into the formulation of

classical differential equations expand the variety and complexity of possibles behavior regimes often allowing a better description of the real phenomenon [107]. In particular, delays play a key role in Biomathematics (population dynamics, infectious diseases, physiology, biotic population, immunology, neural networks and cell kinetics) [10, 37, 91], but also in other realms like Chemistry [65, Ch. 4], Engineering [108], Economics and Finance [82, 122].

As it has been previously indicated, delay differential equations allow describing more complex dynamics than their classical counterpart. This fact is particularly convenient in dealing with modeling using real data where, in addition, it is necessary to perform a rigorous treatment of uncertainty (uncertainty quantification). This randomness usually comes from sampling or simply because of the inherent complexity of the phenomena under study. In this setting, stochastic and random delay differential equations are formulated instead.

On the one hand, stochastic delay differential equations are those in which uncertainty in driven by stochastic processes whose sample path behavior is irregular (typically Brownian motion, or more generally Wiener process, and Poisson process). Their mathematical study requires Itô or Malliavin calculus [129]. Under this approach, uncertainty is limited to specific probabilistic patterns. In the case of considering the Wiener process, then the underlying noise is of Gaussian type. An excellent overview of this approach can be found in [147]. While some recent theoretical and numerical advances using Gaussian and Poisson distributions are reported in the book [83, Ch. 1 and Ch. 10] and in the articles [6, 71, 126, 148, 149], for example. Stochastic delay differential equations have also been successfully applied to model real problems in different settings. For example, mathematical models to describe the dynamics of obesity and alcohol consumption have been proposed in [141] and [142], respectively. The stochastic Navier-Stokes with infinite delay has been recently addressed in [117]. A predator prey stochastic model with delay has been proposed in [32].

On the other hand, random delay differential equations are those in which the random effects are directly manifested in their inputs (coefficients, initial/boundary conditions and/or source term). The sample path behavior of these inputs is regular (e.g. sample path continuous) with respect to time and space [157, p. 97]. The rigorous analysis of this type of differential equations can be conducted mainly by using two approaches, sample path calculus or mean square random calculus [160]. The former approach is strongly based upon the well-behavior (regularity) of the trajectories of the inputs involved in the random differential equations in order to take advantage of

the power of deterministic calculus. In the latter case, results are formulated in the setting of the Hilbert space $(\mathrm{L}^2, \langle\cdot,\cdot\rangle)$ of real random variables on Ω having second-order moment (thus having mean and variance too) endowed with the inner product $\langle U, V\rangle = \mathbb{E}[UV]$, where $\mathbb{E}[\cdot]$ denotes the expectation operator and Ω is the sample space of an underlying complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ [160]. Mean square random calculus has been successfully applied to study random differential equations, see for example [127, 160]. However, to the best of our knowledge, in the context of random delay differential equations only a few theoretical results have been established. Some recently and very interesting contributions focusing on numerical methods instead are [150, 175]. In [175], a sparse grid stochastic Legendre spectral collocation method is proposed to numerically solve linear systems of random differential equations with constant and pantograph delays. While in [150], the authors extend the generalized polynomial chaos method to study nonlinear random delay differential equations by taking advantage of orthogonality properties in the Hilbert space L^2. Nevertheless, there is lack of theoretical results for random delay differential equations, starting from the random autonomous linear differential equation with delay, in the general context of random Lebesgue spaces $(\mathrm{L}^p, \|\cdot\|_p)$, where $\|U\|_p = (\mathbb{E}[|U|^p])^{1/p}$, for $1 \leq p < \infty$, and $\|U\|_\infty = \inf\{C \geq 0 : |U(\omega)| \leq C \text{ for almost every } \omega\}$, for $U : \Omega \to \mathbb{R}$ being a random variable. And this is precisely the aim of this contribution.

Apart from stochastic and random delay differential equations, it must also be mentioned a complementary approach usually referred to as fuzzy delay differential equations, whose uncertainty is driven by particular stochastic processes like the fuzzy Liu process [120].

Finally, it must be pointed out that randomness is directly introduced in the delay instead of coefficients and/or forcing term in order to account for uncertainties associated to the time instant in which relevant factors determining the output of the mathematical model under study take place. Examples in this regard can be found in [72, 105, 118], for example.

The autonomous linear differential equation with discrete time delay $\tau > 0$ is given by

$$\begin{cases} x'(t) &= ax(t) + bx(t-\tau),\ t \geq 0, \\ x(t) &= g(t),\ -\tau \leq t \leq 0, \end{cases} \tag{10.1}$$

where a is the coefficient of the non-delay component, b is the parameter of the delay term, and the function $g(t)$ defined on $[-\tau, 0]$ is the initial condition. If $g \in C^1([-\tau, 0])$, then the unique solution to (10.1) is obtained with the

method of steps and is given by [102, Th. 1],

$$x(t) = \mathrm{e}^{a(t+\tau)}\mathrm{e}_\tau^{b_1,t}g(-\tau) + \int_{-\tau}^{0} \mathrm{e}^{a(t-s)}\mathrm{e}_\tau^{b_1,t-\tau-s}(g'(s) - ag(s))\,\mathrm{d}s, \tag{10.2}$$

where $b_1 = \mathrm{e}^{-a\tau}b$,

$$\mathrm{e}_\tau^{c,t} = \begin{cases} 0, & -\infty < t < -\tau, \\ 1, & -\tau \le t < 0, \\ 1 + c\dfrac{t}{1!}, & 0 \le t < \tau, \\ 1 + c\dfrac{t}{1!} + c^2\dfrac{(t-\tau)^2}{2!}, & \tau \le t < 2\tau, \\ \vdots & \vdots \\ \displaystyle\sum_{k=0}^{n} c^k \frac{(t-(k-1)\tau)^k}{k!}, & (n-1)\tau \le t < n\tau, \end{cases}$$

is the delayed exponential function [102, Def. 1], $c, t \in \mathbb{R}$, $\tau > 0$ and $n = \lfloor t/\tau \rfloor + 1$ (here $\lfloor \cdot \rfloor$ denotes the integer part defined by the so-called floor function).

The randomization of (10.1) consists in assuming that the system depends on an outcome ω of an experiment:

$$\begin{cases} x'(t,\omega) &= a(\omega)x(t,\omega) + b(\omega)x(t-\tau,\omega),\ t \ge 0, \\ x(t,\omega) &= g(t,\omega),\ -\tau \le t \le 0. \end{cases} \tag{10.3}$$

Here, the coefficients $a = a(\omega)$ and $b = b(\omega)$ are random variables, while $g(t) = g(t,\omega)$ is a stochastic process, all of them defined in an underlying complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

The formal solution to (10.3) is obtained after randomization of (10.2):

$$\begin{aligned} x(t,\omega) = {} & \mathrm{e}^{a(\omega)(t+\tau)}\mathrm{e}_\tau^{b_1(\omega),t}g(-\tau,\omega) \\ & + \int_{-\tau}^{0} \mathrm{e}^{a(\omega)(t-s)}\mathrm{e}_\tau^{b_1(\omega),t-\tau-s}(g'(s,\omega) - a(\omega)g(s,\omega))\,\mathrm{d}s, \end{aligned} \tag{10.4}$$

where $b_1(\omega) = \mathrm{e}^{-a(\omega)\tau}b(\omega)$. The stochastic process (10.4) is a solution to (10.3) in the sample path sense, under the assumption that the sample paths of g belong to $C^1([-\tau, 0])$.

In this chapter, we study conditions under which (10.4) is an L^p-solution to (10.3). This kind of delay differential equations appear in Engineering and Control problems [5, 108], for example. When they are applied to real data,

its coefficients a and b and its *preshape* function $g(t)$ need to be calibrated. Since data often involve uncertainty coming from errors measurements, it is more realistic to treat a and b as random variables and $g(t)$ as a stochastic process, as it will be assumed throughout this chapter.

This chapter is organized as follows. In Section 10.2, we state preliminary results on L^p-calculus that are required for the exposition. In Section 10.3, we prove that (10.4) is the unique L^p-solution to (10.3) under certain assumptions. In Section 10.4, we demonstrate that (10.4) tends as $\tau \to 0$ to the solution to (10.3) with $\tau = 0$, in the space L^p.

10.2 Preliminary results on L^p-calculus

In this section, we state some preliminary results on L^p-calculus that will be required in the coming sections.

Proposition 10.1 (Chain Rule Theorem) *Let $\{X(t) : t \in [a,b]\}$ be any stochastic process. Let f be a deterministic C^1 function on an open set that contains $X([a,b])$. Fix $1 \leq p < \infty$. Let $t \in [a,b]$ such that:*

(i) X is L^{2p}-differentiable at t,

(ii) X is path continuous on $[a,b]$,

(iii) there exist $r > 2p$ and $\delta > 0$ such that $\sup_{s\in[-\delta,\delta]} \mathbb{E}[|f'(X(t+s))|^r] < \infty$.

Then $f \circ X$ is L^p-differentiable at t and $(f \circ X)'(t) = f'(X(t))X'(t)$.

The proof of Proposition 10.1 is analogous to [167, Th. 3.19], but not restricted to mean square and mean fourth calculus. In the proof, instead of applying Hölder's inequality as $\|UV\|_2 \leq \|U\|_4 \|V\|_4$, one uses the more general version $\|UV\|_p \leq \|U\|_{2p} \|V\|_{2p}$ (here U and V represent random variables).

Lemma 10.2 *Let $Y_1(t,s)$, $Y_2(t,s)$ and $Y_3(t,s)$ be three stochastic processes and fix $1 \leq p < \infty$. If Y_1 and Y_2 are L^q-continuous for all $1 \leq q < \infty$, and Y_3 is $\mathrm{L}^{p+\eta}$-continuous for certain $\eta > 0$, then the product process $Y_1Y_2Y_3$ is L^p-continuous.*

On the other hand, if Y_1 and Y_2 are L^∞-continuous, and Y_3 is L^p-continuous, then the product process $Y_1Y_2Y_3$ is L^p-continuous.

Proof. Suppose that Y_1 and Y_2 are L^q-continuous for all $1 \leq q < \infty$, and Y_3 is $\mathrm{L}^{p+\eta}$-continuous. Notice that Y_1Y_2 is L^q-continuous for all $1 \leq q < \infty$. Indeed, if $t_n \to t$ and $s_n \to s$ as $n \to \infty$, then, by the triangular and Hölder's inequalities,

$$\begin{aligned}
&\|Y_1(t_n,s_n)Y_2(t_n,s_n) - Y_1(t,s)Y_2(t,s)\|_q \\
&\leq \|(Y_1(t_n,s_n) - Y_1(t,s))Y_2(t_n,s_n)\|_q + \|Y_1(t,s)(Y_2(t,s) - Y_2(t_n,s_n))\|_q \\
&\leq \|Y_1(t_n,s_n) - Y_1(t,s)\|_{2q}\|Y_2(t_n,s_n)\|_{2q} \\
&+\|Y_1(t,s)\|_{2q}\|Y_2(t,s) - Y_2(t_n,s_n)\|_{2q} \overset{n\to\infty}{\longrightarrow} 0\,. \qquad (10.5)
\end{aligned}$$

Now,

$$\begin{aligned}
&\|Y_1(t_n,s_n)Y_2(t_n,s_n)Y_3(t_n,s_n) - Y_1(t,s)Y_2(t,s)Y_3(t,s)\|_p \\
&\leq \|Y_1(t_n,s_n)Y_2(t_n,s_n)(Y_3(t_n,s_n) - Y_3(t,s))\|_p \\
&+\|(Y_1(t_n,s_n)Y_2(t_n,s_n) - Y_1(t,s)Y_2(t,s))Y_3(t,s)\|_p \\
&\leq \|Y_1(t_n,s_n)Y_2(t_n,s_n)\|_q\|Y_3(t_n,s_n) - Y_3(t,s)\|_{p+\eta} \\
&+\|Y_1(t_n,s_n)Y_2(t_n,s_n) - Y_1(t,s)Y_2(t,s)\|_q\|Y_3(t,s)\|_{p+\eta} \overset{n\to\infty}{\longrightarrow} 0, \qquad (10.6)
\end{aligned}$$

where $q = \frac{p(p+\eta)}{\eta}$ has been chosen to apply Hölder's inequality (note that $\frac{1}{p} = \frac{1}{p+\eta} + \frac{1}{q}$). This proves the L^p-continuity of $Y_1Y_2Y_3$.
Suppose that Y_1 and Y_2 are L^∞-continuous, and Y_3 is L^p-continuous. Then Y_1Y_2 is L^∞-continuous, by (10.5) with $q = \infty$. Statement (10.6) holds with $q = \infty$ and p in lieu of $p + \eta$. This demonstrates the L^p-continuity of $Y_1Y_2Y_3$. □

Lemma 10.3 *Let $Y_1(t)$, $Y_2(t)$ and $Y_3(t)$ be three stochastic processes, and $1 \leq p < \infty$. If Y_1 and Y_2 are L^q-differentiable for all $1 \leq q < \infty$, and Y_3 is $\mathrm{L}^{p+\eta}$-differentiable for certain $\eta > 0$, then the product process $Y_1Y_2Y_3$ is L^p-differentiable and $\frac{\mathrm{d}}{\mathrm{d}t}(Y_1(t)Y_2(t)Y_3(t)) = Y_1'(t)Y_2(t)Y_3(t) + Y_1(t)Y_2'(t)Y_3(t) + Y_1(t)Y_2(t)Y_3'(t)$.*

Additionally, if Y_1 and Y_2 are assumed to be L^∞-differentiable, and Y_3 is L^p-differentiable, then $Y_1Y_2Y_3$ is L^p-differentiable, with $\frac{\mathrm{d}}{\mathrm{d}t}(Y_1(t)Y_2(t)Y_3(t)) = Y_1'(t)Y_2(t)Y_3(t) + Y_1(t)Y_2'(t)Y_3(t) + Y_1(t)Y_2(t)Y_3'(t)$.

The proof of this lemma follows the same reasoning as Lemma 10.2, but working with incremental quotients instead. Similar reasonings are also given in [160, p. 96 (4)], [167, Lemma 3.14]. We omit the details.

Proposition 10.4 (L^p-differentiation under L^p-Riemann integral sign) *Let $F(t,s)$ be a stochastic process on $[a,b]\times[c,d]$. Fix $1\le p<\infty$. Suppose that $F(t,\cdot)$ is L^p-continuous on $[c,d]$, for each $t\in[a,b]$, and that there exists the L^p-partial derivative $\frac{\partial F}{\partial t}(t,s)$ for all $(t,s)\in[a,b]\times[c,d]$, which is L^p-continuous on $[a,b]\times[c,d]$. Let $G(t)=\int_c^d F(t,s)\,\mathrm{d}s$ (the integral is understood as an L^p-Riemann integral). Then G is L^p-differentiable on $[a,b]$ and $G'(t)=\int_c^d \frac{\partial F}{\partial t}(t,s)\,\mathrm{d}s$.*

Proof. We have, for $h\neq 0$,

$$\begin{aligned}
&\left\|\frac{G(t+h)-G(t)}{h}-\int_c^d\frac{\partial F}{\partial t}(t,s)\,\mathrm{d}s\right\|_p\\
&=\left\|\int_c^d\left(\frac{F(t+h,s)-F(t,s)}{h}-\frac{\partial F}{\partial t}(t,s)\right)\mathrm{d}s\right\|_p\\
&\le\int_c^d\left\|\frac{F(t+h,s)-F(t,s)}{h}-\frac{\partial F}{\partial t}(t,s)\right\|_p\mathrm{d}s, \qquad (10.7)
\end{aligned}$$

where the last inequality comes from [160, p. 102] in the general setting of L^p-calculus. We know that

$$\lim_{h\to 0}\left\|\frac{F(t+h,s)-F(t,s)}{h}-\frac{\partial F}{\partial t}(t,s)\right\|_p=0,$$

by definition of L^p-partial derivative. We bound $\|\frac{F(t+h,s)-F(t,s)}{h}-\frac{\partial F}{\partial t}(t,s)\|_p$ in order to apply the Dominated Convergence Theorem in (10.7).
On the one hand,

$$\left\|\frac{\partial F}{\partial t}(t,s)\right\|_p\le M, \qquad (10.8)$$

for all $(t,s)\in[a,b]\times[c,d]$, by L^p-continuity of $\frac{\partial F}{\partial t}(t,s)$. On the other hand, by Barrow's rule [160, p. 104], an inequality from [160, p. 102], and (10.8),

$$\begin{aligned}
\left\|\frac{F(t+h,s)-F(t,s)}{h}\right\|_p&=\frac{1}{|h|}\left\|\int_t^{t+h}\frac{\partial F}{\partial t}(t',s)\,\mathrm{d}t'\right\|_p\\
&\le\frac{1}{|h|}\left|\int_t^{t+h}\left\|\frac{\partial F}{\partial t}(t',s)\right\|_p\mathrm{d}t'\right|\le M. \qquad (10.9)
\end{aligned}$$

By the triangular inequality, (10.8) and (10.9),

$$\begin{aligned} &\left\| \frac{F(t+h,s)-F(t,s)}{h} - \frac{\partial F}{\partial t}(t,s) \right\|_p \\ &\leq \left\| \frac{F(t+h,s)-F(t,s)}{h} \right\|_p + \left\| \frac{\partial F}{\partial t}(t,s) \right\|_p \leq 2M, \end{aligned}$$

so the use of the Dominated Convergence Theorem to conclude that (10.7) tends to 0 as $h \to 0$ is justified.

□

10.3 L^p-solution to the random autonomous linear differential equation with discrete delay

In this section, we solve (10.3) in the L^p-sense. We will establish its uniqueness of solution, and we will prove that (10.4) is an L^p-solution under certain conditions (the integral from (10.4) will be understood as an L^p-Riemann integral).

Theorem 10.5 (Uniqueness) *The stochastic system (10.3) has at most one L^p-solution, for $1 \leq p < \infty$.*

Proof. Suppose that $x(t)$ and $y(t)$ are two L^p-solutions to (10.3). Let $z(t) = x(t) - y(t)$, which satisfies the random differential equation problem with delay

$$\begin{cases} z'(t,\omega) = a(\omega)z(t,\omega) + b(\omega)z(t-\tau,\omega), \ t \geq 0, \\ z(t,\omega) = 0, \ -\tau \leq t \leq 0. \end{cases}$$

If $t \in [0,\tau]$, then $t - \tau \in [-\tau, 0]$, therefore $z(t-\tau) = 0$. Thus, $z(t)$ satisfies a random differential equation problem with no delay:

$$\begin{cases} z'(t,\omega) = a(\omega)z(t,\omega), \ t \in [0,\tau], \\ z(0,\omega) = 0. \end{cases} \tag{10.10}$$

In [161], it was proved that any L^p-solution to a random initial value problem has a product measurable representative which is an absolutely continuous solution in the sample path sense. Since the sample path solution to (10.10) must be 0 (from the deterministic theory), we conclude that $z(t) = 0$, as wanted.

□

Proposition 10.6 (L^p-derivative of the delayed exponential function)

Consider the stochastic system with discrete delay

$$\begin{cases} x'(t,\omega) = c(\omega)x(t-\tau,\omega),\ t \geq 0, \\ x(t,\omega) = 1,\ -\tau \leq t \leq 0, \end{cases} \tag{10.11}$$

where $c(\omega)$ is a random variable.

If c has centered absolute moments of any order, then $\mathrm{e}_\tau^{c,t}$ is the unique L^p-solution to (10.11), for all $1 \leq p < \infty$.

On the other hand, if c is bounded, then $\mathrm{e}_\tau^{c,t}$ is the unique L^∞-solution to (10.11).

Proof. Suppose that c has centered absolute moments of any order, and let $x(t) = \mathrm{e}_\tau^{c,t}$. Fix $t_0 \geq 0$. We want to prove that x is L^p-differentiable at t_0, for all $1 \leq p < \infty$, with $x'(t_0) = cx(t_0 - \tau)$.
For $n = \lfloor t_0/\tau \rfloor + 1$, t_0 belongs to $[(n-1)\tau, n\tau)$. We distinguish two cases: $(n-1)\tau < t_0 < n\tau$ and $t_0 = (n-1)\tau$.
In the former case, $\mathrm{e}_\tau^{c,t} = \sum_{k=0}^{n} c^k \frac{(t-(k-1)\tau)^k}{k!}$ for all $t \in ((n-1)\tau, n\tau)$, which is a neighborhood of t_0. Each addend $c^k \frac{(t-(k-1)\tau)^k}{k!}$ is L^p-differentiable, with derivative $\frac{\mathrm{d}}{\mathrm{d}t}\{c^k \frac{(t-(k-1)\tau)^k}{k!}\} = c^k \frac{(t-(k-1)\tau)^{k-1}}{(k-1)!}$:

$$\left\| c^k \frac{\frac{(t+h-(k-1)\tau)^k}{k!} - \frac{(t-(k-1)\tau)^k}{k!}}{h} - c^k \frac{(t-(k-1)\tau)^{k-1}}{(k-1)!} \right\|_p$$
$$= \|c^k\|_p \left| \frac{\frac{(t+h-(k-1)\tau)^k}{k!} - \frac{(t-(k-1)\tau)^k}{k!}}{h} - \frac{(t-(k-1)\tau)^{k-1}}{(k-1)!} \right| \xrightarrow{h\to 0} 0,$$

since c has centered absolute moments of any order and by the classical derivative of $\frac{(t-(k-1)\tau)^k}{k!}$. Then $x(t) = \mathrm{e}_\tau^{c,t}$ is L^p-differentiable on $((n-1)\tau, n\tau)$ and $x'(t) = \sum_{k=1}^{n} c^k \frac{(t-(k-1)\tau)^{k-1}}{(k-1)!} = c\sum_{k=0}^{n-1} c^k \frac{(t-k\tau)^k}{k!} = cx(t-\tau)$. Notice that if c were bounded, the limit computed as $h \to 0$ holds with $p = \infty$, so $x(t)$ is L^∞-differentiable on $((n-1)\tau, n\tau)$.
In the latter case $t_0 = (n-1)\tau$, we need to compute the left and right derivatives of $x(t) = \mathrm{e}_\tau^{c,t}$ at t_0, and check that both are equal to $cx(t_0 - \tau) =$

$c\sum_{k=0}^{n-1} c^k \frac{(t_0-k\tau)^k}{k!} = \sum_{k=1}^{n} c^k \frac{(t_0-(k-1)\tau)^{k-1}}{(k-1)!}$. On the one hand, for $h>0$,

$$\begin{aligned}
&\left\| \frac{x(t_0+h)-x(t_0)}{h} - \sum_{k=1}^{n} c^k \frac{(t_0-(k-1)\tau)^{k-1}}{(k-1)!} \right\|_p \\
&\leq \sum_{k=1}^{n} \left\| c^k \frac{\frac{(t_0+h-(k-1)\tau)^k}{k!} - \frac{(t_0-(k-1)\tau)^k}{k!}}{h} - c^k \frac{(t_0-(k-1)\tau)^{k-1}}{(k-1)!} \right\|_p \\
&= \sum_{k=1}^{n} \|c^k\|_p \left| \frac{\frac{(t_0+h-(k-1)\tau)^k}{k!} - \frac{(t_0-(k-1)\tau)^k}{k!}}{h} - \frac{(t_0-(k-1)\tau)^{k-1}}{(k-1)!} \right| \overset{h\to 0^+}{\longrightarrow} 0,
\end{aligned}$$

by the classical derivative of $\frac{(t-(k-1)\tau)^k}{k!}$ and the boundedness of the absolute moments of c. On the other hand, for $h<0$,

$$\begin{aligned}
&\left\| \frac{x(t_0+h)-x(t_0)}{h} - \sum_{k=1}^{n} c^k \frac{(t_0-(k-1)\tau)^{k-1}}{(k-1)!} \right\|_p \\
&= \left\| \sum_{k=1}^{n-1} c^k \frac{\frac{(t_0+h-(k-1)\tau)^k}{k!} - \frac{(t_0-(k-1)\tau)^k}{k!}}{h} - \sum_{k=1}^{n-1} c^k \frac{(t_0-(k-1)\tau)^{k-1}}{(k-1)!} \right\|_p \\
&\leq \sum_{k=1}^{n-1} \|c^k\|_p \left| \frac{\frac{(t_0+h-(k-1)\tau)^k}{k!} - \frac{(t_0-(k-1)\tau)^k}{k!}}{h} - \frac{(t_0-(k-1)\tau)^{k-1}}{(k-1)!} \right| \overset{h\to 0^-}{\longrightarrow} 0.
\end{aligned}$$

This proves that $x(t) = \mathrm{e}_\tau^{c\cdot t}$ is L^p-differentiable at t_0, with $x'(t_0) = cx(t-t_0)$. Again, note that if c were bounded, the limits computed as $h\to 0$ hold with $p=\infty$, so $x(t)$ is L^∞-differentiable at t_0.

□

In what follows, we denote the moment-generating function of a random variable a as $\phi_a(\zeta) = \mathbb{E}[\mathrm{e}^{a\zeta}]$, $\zeta\in\mathbb{R}$.

Theorem 10.7 (Existence and uniqueness) *Fix $1\leq p<\infty$. Suppose that $\phi_a(\zeta)<\infty$ for all $\zeta\in\mathbb{R}$, b has centered absolute moments of any order, and g belongs to $C^1([-\tau,0])$ in the $\mathrm{L}^{p+\eta}$-sense, for certain $\eta>0$. Then the stochastic process $x(t)$ defined by (10.4) is the unique L^p-solution to (10.3).*

Proof. Let us see that $b_1 = \mathrm{e}^{-a\tau}b$ has centered absolute moments of any order. For $m\geq 0$ and by Cauchy-Schwarz inequality,

$$\begin{aligned}
\mathbb{E}[|b_1|^m] &= \mathbb{E}[\mathrm{e}^{-ma\tau}|b|^m] \leq \left(\mathbb{E}[\mathrm{e}^{-2ma\tau}]\right)^{\frac{1}{2}} \left(\mathbb{E}[b^{2m}]\right)^{\frac{1}{2}} \\
&= \left(\phi_a(-2m\tau)\right)^{\frac{1}{2}} \left(\mathbb{E}[b^{2m}]\right)^{\frac{1}{2}} < \infty.
\end{aligned}$$

By Proposition 10.6, $\mathrm{e}_\tau^{b_1,t}$ is L^q-differentiable, for each $1 \leq q < \infty$, and $\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}_\tau^{b_1,t} = b_1\mathrm{e}_\tau^{b_1,t-\tau}$.

Consider the stochastic process e^{at}. Let us check the conditions of Proposition 10.1 (Chain Rule Theorem) with $f(x) = \mathrm{e}^x$, $X(t) = at$ and any $1 \leq q < \infty$. Notice that, from $\phi_a(\zeta) < \infty$ for all $\zeta \in \mathbb{R}$, we know that a has centered absolute moments of any order and that ϕ_a is an analytic real function with $\phi_a(\zeta) = \sum_{n=0}^{\infty} \frac{\zeta^n}{n!}\mathbb{E}[a^n]$. These facts give conditions (i), (ii) and (iii) from Proposition 10.1. Indeed, (i) holds because a has centered absolute moments of any order, so in particular $a \in \mathrm{L}^{2q}$, which gives the L^{2q}-differentiability of at; (ii) is clear since at is path-continuous; and (iii) requires, for each $t \geq 0$, an $r > 2q$ and a $\delta > 0$ such that $\sup_{s\in[-\delta,\delta]} \phi_a(r(t+s)) < \infty$, but this is clear by continuity of ϕ_a on $\mathbb{R}$ and its boundedness on any compact set. The conclusion is that e^{at} is L^q-differentiable, for each $1 \leq q < \infty$, and $\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{at} = a\mathrm{e}^{at}$.

We apply Lemma 10.3 with $Y_1(t) = \mathrm{e}^{a(t+\tau)}$, $Y_2(t) = \mathrm{e}_\tau^{b_1,t}$ and $Y_3(t) = g(-\tau)$. We saw that Y_1 and Y_2 are L^q-differentiable for all $1 \leq q < \infty$, and $Y_3(t) \in \mathrm{L}^{p+\eta}$ by hypothesis, therefore the product process $Y_1(t)Y_2(t)Y_3(t) = \mathrm{e}^{a(t+\tau)}\mathrm{e}_\tau^{b_1,t}g(-\tau)$ is L^p-differentiable and

$$\frac{\mathrm{d}}{\mathrm{d}t}\left\{\mathrm{e}^{a(t+\tau)}\mathrm{e}_\tau^{b_1,t}g(-\tau)\right\} = \left\{a\mathrm{e}^{a(t+\tau)}\mathrm{e}_\tau^{b_1,t} + \mathrm{e}^{a(t+\tau)}b_1\mathrm{e}_\tau^{b_1,t-\tau}\right\}g(-\tau). \quad (10.12)$$

Let $Y_1(t,s) = \mathrm{e}^{a(t-s)}$, $Y_2(t,s) = \mathrm{e}_\tau^{b_1,t-\tau-s}$ and $Y_3(t,s) = g'(s) - ag(s)$. Set $F(t,s) = Y_1(t,s)Y_2(t,s)Y_3(t,s)$. On the one hand, since e^{at} and $\mathrm{e}_\tau^{b_1,t}$ are L^q-continuous processes, for each $1 \leq q < \infty$, $Y_1(t,s)$ and $Y_2(t,s)$ are L^q-continuous at (t,s). On the other hand, from $g \in C^1([-\tau,0])$ in the $\mathrm{L}^{p+\eta}$-sense, a having absolute moments of any order, and Hölder's inequality, we derive that $Y_3(t,s)$ is $\mathrm{L}^{p+\mu}$-continuous, for $0 < \mu < \eta$. By Lemma 10.2, $F(t,s)$ is L^p-continuous at (t,s).

Fixed s, let $Y_1(t) = \mathrm{e}^{a(t-s)}$, $Y_2(t) = \mathrm{e}_\tau^{b_1,t-\tau-s}$ and $Y_3(t) = g'(s) - ag(s)$. We know that $Y_1(t)$ and $Y_2(t)$ are L^q-differentiable, for each $1 \leq q < \infty$. Also, from $g \in C^1([-\tau,0])$ in the $\mathrm{L}^{p+\eta}$-sense, a having absolute moments of any order, and Hölder's inequality, the random variable $Y_3(t)$ belongs to $\mathrm{L}^{p+\mu}$, for all $0 < \mu < \eta$. By Lemma 10.3, $F(\cdot,s)$ is L^p-differentiable at t, with

$$\frac{\partial F}{\partial t}(t,s) = \left\{a\mathrm{e}^{a(t-s)}\mathrm{e}_\tau^{b_1,t-\tau-s} + \mathrm{e}^{a(t-s)}b_1\mathrm{e}_\tau^{b_1,t-2\tau-s}\right\}\left(g'(s) - ag(s)\right). \quad (10.13)$$

Let us see that $\frac{\partial F}{\partial t}(t,s)$ is L^p-continuous at (t,s). Since a has centered absolute moments of any order and $\mathrm{e}^{a(t-s)}$ is L^q-continuous at (t,s), for each $1 \leq q < \infty$, we derive that $a\mathrm{e}^{a(t-s)}$ is L^q-continuous at (t,s), for each $1 \leq q < \infty$, by Hölder's inequality. We have that $Y_1(t,s) = a\mathrm{e}^{a(t-s)}$ and $Y_2(t,s) = \mathrm{e}_\tau^{b_1,t-\tau-s}$ are L^q-continuous at (t,s), for each $1 \leq q < \infty$, while $Y_3(t,s) = g'(s) - ag(s)$ is $\mathrm{L}^{p+\mu}$-continuous, for $0 < \mu < \eta$. By Lemma 10.2, $a\mathrm{e}^{a(t-s)}\mathrm{e}_\tau^{b_1,t-\tau-s}(g'(s) -$

$ag(s))$ is L^p-continuous at each (t,s). Analogously, $\mathrm{e}^{a(t-s)}b_1\mathrm{e}_\tau^{b_1,t-2\tau-s}(g'(s) - ag(s))$ is L^p-continuous at (t,s). Therefore, $\frac{\partial F}{\partial t}(t,s)$ is L^p-continuous at (t,s). Set $G(t) = \int_{-\tau}^0 F(t,s)\,\mathrm{d}s$. By Proposition 10.4 and (10.13), the process G is L^p-differentiable and

$$\begin{aligned} G'(t) &= \int_{-\tau}^0 \frac{\partial F}{\partial t}(t,s)\,\mathrm{d}s \\ &= \int_{-\tau}^0 \left\{ a\mathrm{e}^{a(t-s)}\mathrm{e}_\tau^{b_1,t-\tau-s} + \mathrm{e}^{a(t-s)}b_1\mathrm{e}_\tau^{b_1,t-2\tau-s} \right\} (g'(s) - ag(s))\,\mathrm{d}s. \end{aligned} \tag{10.14}$$

By combining (10.12) and (10.14) and taking into account expression (10.4) for $x(t)$, we derive that $x(t)$ is L^p-differentiable and $x'(t) = ax(t) + bx(t-\tau)$, after simple operations.

Finally, for the initial condition, we put $t \in [-\tau, 0]$ into (10.4). For the first addend, note that $\mathrm{e}_\tau^{b_1,t} = 1$ by definition of exponential delay function. For the second addend, we need to work more. Let $s \in [-\tau, 0]$. If $-\tau \le t < s \le 0$, then $t - \tau \le t - \tau - s < -\tau$, so $\mathrm{e}_\tau^{b_1,t-\tau-s} = 0$. If $-\tau \le s \le t \le 0$, then $-\tau \le t - \tau - s \le t$, therefore $\mathrm{e}_\tau^{b_1,t-\tau-s} = 1$. Thus, $x(t) = \mathrm{e}^{a(t+\tau)}g(-\tau) + \int_{-\tau}^t \mathrm{e}^{a(t-s)}(g'(s) - ag(s))\,\mathrm{d}s$. Fixed t, let $Y_1(s) = 1$, $Y_2(s) = \mathrm{e}^{a(t-s)}$ and $Y_3(s) = g(s)$. By Lemma 10.3, the product process $Y_1(s)Y_2(s)Y_3(s) = \mathrm{e}^{a(t-s)}g(s)$ is L^p-differentiable, with $\frac{\mathrm{d}}{\mathrm{d}s}(\mathrm{e}^{a(t-s)}g(s)) = \mathrm{e}^{a(t-s)}(g'(s) - ag(s))$, which is L^p-continuous. By Barrow's rule for L^p-calculus, see [160, p. 104] (in the setting of mean square calculus),

$$\begin{aligned} x(t) &= \mathrm{e}^{a(t+\tau)}g(-\tau) + \int_{-\tau}^t \mathrm{e}^{a(t-s)}(g'(s) - ag(s))\,\mathrm{d}s \\ &= \mathrm{e}^{a(t+\tau)}g(-\tau) + \int_{-\tau}^t \frac{\mathrm{d}}{\mathrm{d}t}(\mathrm{e}^{a(t-s)}g(s))\,\mathrm{d}s = g(t). \end{aligned}$$

□

Example 10.8 Consider the delay $\tau = 2$. Set $a \sim \mathrm{Normal}(2,1)$ and $b \sim \mathrm{Gamma}(2,2)$. Let $g(t) = \sin(dt^2)$, where $d \sim \mathrm{Beta}(10,9)$. We know that $\phi_a(\zeta) < \infty$ for all $\zeta \in \mathbb{R}$, and that b has centered absolute moments of any order. On the other hand, $g(t)$ is L^p-differentiable on $\mathbb{R}$, for each $1 \le p < \infty$, as a consequence of Proposition 10.1. Indeed, in the notation of Proposition 10.1, take $f(x) = \sin x$ and $X(t) = dt^2$. Condition (i) holds because d is bounded, so it has absolute moments of any order. Condition (ii) is obvious. Finally, condition (iii) follows since $|f'(x)| = |\cos x| \le 1$. Therefore $g(t)$ is L^p-differentiable, for each $1 \le p < \infty$, and $g'(t) = 2dt\cos(dt^2)$. In fact, with the same reasoning by applying Proposition 10.1 and the product rule differ-

entiation, g is $C^\infty(\mathbb{R})$. The assumptions of Theorem 10.7 are satisfied, so $x(t)$ defined by (10.4) is the unique L^p-solution to (10.3), for each $1 \leq p < \infty$.

To understand the main probabilistic features of $x(t)$ (uncertainty quantification), one may approximate the statistical moments of $x(t)$ (since $x(t) \in \mathrm{L}^p$ for $1 \leq p < \infty$). The main statistics of $x(t)$ are its expectation, $\mathbb{E}[x(t)]$, and its variance, $\mathbb{V}[x(t)]$. Table 10.1 shows the approximation of these two statistics for different times $t \geq 0$, $\mathbb{E}_{\mathrm{MC}}[x(t)]$ and $\mathbb{V}_{\mathrm{MC}}[x(t)]$, by using Monte Carlo simulation with $2,000,000$ realizations. We have executed the Monte Carlo simulation twice. Observe that the approximations agree, although they deteriorate for large t and the second-order moment.

t	0.1	0.2	0.4	1	1.5
$\mathbb{E}_{\mathrm{MC}}[x(t)]$	0.366665	0.852564	2.25884	14.3022	64.3984
$\mathbb{E}_{\mathrm{MC}}[x(t)]$	0.366769	0.852833	2.25973	14.3123	64.4420
$\mathbb{V}_{\mathrm{MC}}[x(t)]$	0.0784624	0.411983	3.06961	318.432	25090.4
$\mathbb{V}_{\mathrm{MC}}[x(t)]$	0.0784137	0.412158	3.07703	321.213	25008.9

Table 10.1: Approximations of $\mathbb{E}[x(t)]$ and $\mathbb{V}[x(t)]$ with Monte Carlo simulation ($2,000,000$ realizations).

Theorem 10.9 (Existence and uniqueness) *Fix $1 \leq p < \infty$. Suppose that a and b are bounded random variables, and g belongs to $C^1([-\tau, 0])$ in the L^p-sense. Then the stochastic process $x(t)$ defined by (10.4) is the unique L^p-solution to (10.3).*

Proof. First, note that $b_1 = \mathrm{e}^{-a\tau} b$ is bounded, from the assumed boundedness of both a and b. By Proposition 10.6, $\mathrm{e}_\tau^{b_1,t}$ is L^∞-differentiable and $\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}_\tau^{b_1,t} = b_1 \mathrm{e}_\tau^{b_1,t-\tau}$.
On the other hand, e^{at} is L^∞-differentiable and $\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{at} = a\mathrm{e}^{at}$. Indeed, given $h \neq 0$, by the deterministic Mean Value Theorem and the boundedness of a,

$$
\begin{aligned}
&\left\| \frac{\mathrm{e}^{a(t+h)} - \mathrm{e}^{at}}{h} - a\mathrm{e}^{at} \right\|_\infty = \left\| \mathrm{e}^{at} \left(\frac{\mathrm{e}^{ah} - 1}{h} - a \right) \right\|_\infty \\
&\leq \mathrm{e}^{\|a\|_\infty t} \left\| \frac{\mathrm{e}^{ah} - 1}{h} - a \right\|_\infty = \mathrm{e}^{\|a\|_\infty t} \left\| a(\mathrm{e}^{a\xi_h} - 1) \right\|_\infty \\
&\leq \|a\|_\infty \mathrm{e}^{\|a\|_\infty t} \|\mathrm{e}^{a\xi_h} - 1\|_\infty = \|a\|_\infty \mathrm{e}^{\|a\|_\infty t} \|a\mathrm{e}^{a\delta_h}\xi_h\|_\infty \\
&\leq \|a\|_\infty^2 \mathrm{e}^{\|a\|_\infty t} \mathrm{e}^{\|a\|_\infty |h|} |h| \xrightarrow{h \to 0} 0,
\end{aligned}
$$

where ξ_h and δ_h are random variables, that arise from applying twice the deterministic Mean Value Theorem, which depend on h and $|\delta_h| < |\xi_h| < |h|$. The rest of the proof is completely analogous to Theorem 10.7, by applying the second part of both Lemma 10.2 and Lemma 10.3.

□

Remark 10.10 The condition of boundedness for a and b in Theorem 10.9 is necessary if we only assume that $g \in C^1([-\tau, 0])$ in the L^p-sense. See [161, Example p. 541], where it is proved that, in order for a random autonomous and homogeneous linear differential equation of first-order to have an L^p-solution for every initial condition in L^p, one needs the random coefficient to be bounded.

10.4 L^p-convergence to a random autonomous linear differential equation when the delay tends to 0

Given a discrete delay $\tau > 0$, we denote the L^p-solution to (10.3) as $x_\tau(t)$, which is given by (10.4) (we make the dependence on τ explicitly). If we put $\tau = 0$ into (10.3), we obtain a random differential equation problem:

$$\begin{cases} x_0'(t,\omega) = (a(\omega) + b(\omega))x_0(t,\omega),\ t \geq 0, \\ x_0(0,\omega) = g(0). \end{cases} \tag{10.15}$$

Under certain conditions that imitate those from Theorem 10.7 and Theorem 10.9, there exists a unique solution to (10.15), see the forthcoming Proposition 10.11 and subsequent Corollary 10.12. Our objective in this section will be to demonstrate that $\lim_{\tau \to 0} x_\tau(t) = x_0(t)$ in L^p.

Proposition 10.11 *Consider the random differential equation problem*

$$\begin{cases} x_0'(t,\omega) = a(\omega)x_0(t,\omega),\ t \geq 0, \\ x_0(0,\omega) = y_0(\omega), \end{cases} \tag{10.16}$$

where $a(\omega)$ and $y_0(\omega)$ are random variables. Fix $1 \leq p < \infty$.

If $\phi_a(\zeta) < \infty$ for all $\zeta \in \mathbb{R}$, and $y_0 \in \mathrm{L}^{p+\eta}$ for certain $\eta > 0$, then the stochastic process $x_0(t) = y_0 \mathrm{e}^{at}$ is the unique L^p-solution to (10.16).

On the other hand, if a is a bounded random variable and $y_0 \in \mathrm{L}^p$, then the stochastic process $x_0(t) = y_0 \mathrm{e}^{at}$ is the unique L^p-solution to (10.16).

Proof. From $\phi_a(\zeta) < \infty$ for all $\zeta \in \mathbb{R}$, we know that a has centered absolute moments of any order and that ϕ_a is an analytic real function. From these facts, the conditions of Proposition 10.1 with $f(x) = \mathrm{e}^x$ and $X(t) = at$ are easy to check. Hence, e^{at} is an L^q-differentiable stochastic process, for each $1 \leq q < \infty$, with $\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{at} = a\mathrm{e}^{at}$. On the other hand, since $y_0 \in \mathrm{L}^{p+\eta}$, we can turn to Lemma 10.3 with $Y_1(t) = 1$, $Y_2(t) = \mathrm{e}^{at}$ and $Y_3(t) = y_0$ to derive that $x_0(t) = y_0\mathrm{e}^{at}$ is L^p-differentiable and $x_0'(t) = ax_0(t)$.
Finally, if a is bounded, then e^{at} is L^∞-differentiable, check the proof of Theorem 10.9. As $y_0 \in \mathrm{L}^p$, we can turn to the second part of Lemma 10.3 with $Y_1(t) = 1$, $Y_2(t) = \mathrm{e}^{at}$ and $Y_3(t) = y_0$ to conclude that $x_0(t) = y_0\mathrm{e}^{at}$ is L^p-differentiable and $x_0'(t) = ax_0(t)$.

□

Corollary 10.12 *Fix $1 \leq p < \infty$. If $\phi_a(\zeta) < \infty$ and $\phi_b(\zeta) < \infty$ for all $\zeta \in \mathbb{R}$, and $g(0) \in \mathrm{L}^{p+\eta}$ for certain $\eta > 0$, then the stochastic process $x_0(t) = g(0)\mathrm{e}^{(a+b)t}$ is the unique L^p-solution to (10.15).*

On the other hand, if a and b are bounded random variables and $g(0) \in \mathrm{L}^p$, then the stochastic process $x_0(t) = g(0)\mathrm{e}^{(a+b)t}$ is the unique L^p-solution to (10.15).

From now on, we try to demonstrate that $\lim_{\tau\to 0} x_\tau(t) = x_0(t)$ in L^p. First, we prove that the delayed exponential function tends to the classical exponential function in a random setting (Lemma 10.14 and Lemma 10.15), from a well-known deterministic inequality (Lemma 10.13).

Lemma 10.13 *[103, Th. A.3] Let $c \in \mathbb{R}$, $T > 0$, $\tau_0 > 0$ and $\alpha = 1 + |c|\mathrm{e}^{\tau_0|c|}$. Then, for all $\tau \in (0, \tau_0]$, $|\mathrm{e}_\tau^{c,t-\tau} - \mathrm{e}^{ct}| \leq \tau\mathrm{e}^{\alpha T|c|}$, for all $t \in [0, T]$.*

Lemma 10.14 *Let c be a bounded random variable, $T > 0$ and $\tau_0 > 0$. Set $k \geq \|c\|_\infty$. Then, $|\mathrm{e}_\tau^{c(\omega),t-\tau} - e^{c(\omega)t}| \leq C_{T,\tau_0,k} \cdot \tau$, for almost every ω, for all $t \in [0, T]$ and $\tau \in (0, \tau_0]$, for some real constant $C_{T,\tau_0,k} > 0$ that only depends on T, τ_0 and k.*

Proof. By Lemma 10.13, for all $\tau \in (0, \tau_0]$, $|\mathrm{e}_\tau^{c(\omega),t-\tau} - \mathrm{e}^{c(\omega)t}| \leq \tau\mathrm{e}^{\alpha(\omega)T|c(\omega)|}$, for all ω and $t \in [0, T]$. Note that $\alpha(\omega) = 1 + |c(\omega)|\mathrm{e}^{\tau_0|c(\omega)|} \leq 1 + k\mathrm{e}^{\tau_0 k}$. Then $|\mathrm{e}_\tau^{c(\omega),t-\tau} - \mathrm{e}^{c(\omega)t}| \leq \tau\mathrm{e}^{(1+k\mathrm{e}^{\tau_0 k})Tk}$. Now simply set $C_{T,\tau_0,k} = \mathrm{e}^{(1+k\mathrm{e}^{\tau_0 k})Tk}$, and we are done.

□

Lemma 10.15 *Let c be a bounded random variable, $T > 0$ and $\tau_0 > 0$. Set $k \geq \|c\|_\infty$. Then, $|\mathrm{e}_\tau^{c(\omega),t} - e^{c(\omega)t}| \leq C_{T,\tau_0,k} \cdot \tau$, for almost every ω, for all $t \in [-\tau, T]$ and $\tau \in (0, \tau_0]$, for some real constant $C_{T,\tau_0,k} > 0$ that only depends on T, τ_0 and k.*

Proof. By Lemma 10.14, $|\mathrm{e}_\tau^{c(\omega),t} - e^{c(\omega)(t+\tau)}| \leq C'_{T,\tau_0,k} \cdot \tau$, for almost every ω, for all $t \in [-\tau, T]$ and $\tau \in (0, \tau_0]$. By the triangular inequality, $|\mathrm{e}_\tau^{c(\omega),t} - \mathrm{e}^{c(\omega)t}| \leq |\mathrm{e}_\tau^{c(\omega),t} - e^{c(\omega)(t+\tau)}| + |e^{c(\omega)(t+\tau)} - \mathrm{e}^{c(\omega)t}|$. The former addend is bounded by $C'_{T,\tau_0,k} \cdot \tau$, as previously justified. The latter addend is bounded via the deterministic Mean Value Theorem: $|e^{c(\omega)(t+\tau)} - \mathrm{e}^{c(\omega)t}| = \mathrm{e}^{\xi_{\omega,t,\tau}} |c(\omega)| \tau \leq \mathrm{e}^{k(T+\tau_0)} k\tau$, where $\xi_{\omega,t,\tau}$ is between ct and $c(t+\tau)$. Let $C_{T,\tau_0,k} = C'_{T,\tau_0,k} + \mathrm{e}^{k(T+\tau_0)} k$, and we are done.

□

Theorem 10.16 *Fix $1 \leq p < \infty$. Let a and b be bounded random variables and let g be a stochastic process that belongs to $C^1([-\tau, 0])$ in the L^p-sense. Then, $\lim_{\tau \to 0} x_\tau(t) = x_0(t)$ in L^p, uniformly on $[0, T]$, for each $T > 0$.*

Proof. Fix $[0, T]$. Let $\tau_0 = 1$, and $k \geq \|b_1\|_\infty = \|\mathrm{e}^{-a\tau} b\|_\infty$ for all $\tau \in (0, 1]$. Recall that e^{at} is L^∞-continuous (check the proof of Theorem 10.9), therefore $\lim_{\tau \to 0} \mathrm{e}^{a(t+\tau)} = \mathrm{e}^{at}$ in L^∞, uniformly on $[0, T]$. We also have $\lim_{\tau \to 0} g(-\tau) = g(0)$ in L^p, by hypothesis. By Lemma 10.15, $|\mathrm{e}_\tau^{b_1(\omega),t} - e^{b_1(\omega)t}| \leq C_{T,k} \cdot \tau$, so $\|\mathrm{e}_\tau^{b_1,t} - e^{b_1 t}\|_\infty \leq C_{T,k} \cdot \tau$. Since $\lim_{\tau \to 0} \mathrm{e}^{b_1 t} = \mathrm{e}^{bt}$ in L^∞ uniformly on $[0, T]$, we derive that $\lim_{\tau \to 0} \mathrm{e}_\tau^{b_1,t} = \mathrm{e}^{bt}$ in L^∞ uniformly on $[0, T]$. We conclude that

$$\lim_{\tau \to 0} \mathrm{e}^{a(t+\tau)} \mathrm{e}_\tau^{b_1,t} g(-\tau) = g(0) \mathrm{e}^{(a+b)t} = x_0(t) \tag{10.17}$$

in L^p, with uniform convergence on $[0, T]$.
Now, for $\tau \in (0, 1]$ and $t \in [0, T]$,

$$\begin{aligned} & \left\| \int_{-\tau}^{0} \mathrm{e}^{a(t-s)} \mathrm{e}_\tau^{b_1,t-\tau-s} (g'(s) - ag(s)) \, \mathrm{d}s \right\|_p \\ \leq & \int_{-\tau}^{0} \left\| \mathrm{e}^{a(t-s)} \mathrm{e}_\tau^{b_1,t-\tau-s} (g'(s) - ag(s)) \right\|_p \mathrm{d}s \\ \leq & \int_{-\tau}^{0} \left\| \mathrm{e}^{a(t-s)} \right\|_\infty \left\| \mathrm{e}_\tau^{b_1,t-\tau-s} \right\|_\infty \left\| g'(s) - ag(s) \right\|_p \mathrm{d}s. \end{aligned} \tag{10.18}$$

We bound the three terms inside the integral from (10.18). First, $\|\mathrm{e}^{a(t-s)}\|_p \leq \mathrm{e}^{\|a\|_\infty (T+1)}$. Secondly, $\|g'(s) - ag(s)\|_p \leq \|g'(s)\|_p + \|a\|_\infty \|g(s)\|_p \leq C_{\|a\|_\infty, g}$, where $C_{\|a\|_\infty, g} > 0$ is a constant. Finally, we bound $\|\mathrm{e}_\tau^{b_1,t-\tau-s}\|_\infty$. We have

$t \geq 0$ and $s \in [-\tau, 0]$. Then $t - s \in [-\tau, \infty)$. If $t - s \in [-\tau, 0]$, then $t - s - \tau \in [-2\tau, -\tau]$, so $\mathrm{e}_{\tau}^{b_1, t-\tau-s} = 0$ by definition of delayed exponential function. Otherwise, if $t - s \geq 0$, then Lemma 10.14 applies: $\left\| e_{\tau}^{b_1, t-\tau-s} - \mathrm{e}^{b_1(t-s)} \right\|_{\infty} \leq C_{T,k} \cdot \tau$, for $t \in [0, T]$, $s \in [-\tau, 0]$ and $\tau \in (0, 1]$. Since $\left\| \mathrm{e}^{b_1(t-s)} \right\|_{\infty} \leq \mathrm{e}^{k(T+1)}$, we conclude that $\left\| e_{\tau}^{b_1, t-\tau-s} \right\|_{\infty} \leq C_{T,k} \cdot \tau + \mathrm{e}^{k(T+1)}$, by the triangular inequality. Thus, all terms inside the integral from (10.18) are bounded for $t \in [0, T]$ and $s \in [-\tau, 0]$, therefore

$$\lim_{\tau \to 0} \int_{-\tau}^{0} \mathrm{e}^{a(t-s)} \mathrm{e}_{\tau}^{b_1, t-\tau-s} (g'(s) - a g(s)) \, \mathrm{d}s = 0 \tag{10.19}$$

in L^p, uniformly on $[0, T]$.
By combining both (10.17) and (10.19), we conclude that $\lim_{\tau \to 0} x_{\tau}(t) = x_0(t)$ in L^p, uniformly on $[0, T]$.

□

10.5 Conclusions

In this chapter, we have addressed the analysis of the random autonomous linear differential equation with discrete delay. The coefficients have been assumed to be random variables, while the initial condition has been taken as a stochastic process. Although the sample-path approach is the easiest extension of the deterministic results to a random framework, an L^p-random calculus approach is usually the most appropriate method. Uncertainty quantification for stochastic systems requires the computation or approximation of the statistical moments of the solution stochastic process (for instance, via Monte Carlo simulation). Only if we know that the solution process belongs to L^p, we guarantee that the computation or approximation of its statistical moments makes sense. This chapter establishes general conditions under which the random autonomous linear differential equation with discrete delay has a unique L^p-solution. An analysis of L^p-convergence when the delay tends to 0 has also been performed in detail. Our methodology could be extended to other random differential equations with some sort of delay. This will be done in future contributions.

Acknowledgements

This work has been supported by the Spanish Ministerio de Economía y Competitividad grant MTM2017–89664–P. I acknowledge the doctorate scholarship granted by Programa de Ayudas de Investigación y Desarrollo (PAID), Universitat Politècnica de València.

The main results of this chapter have been published in [17].

Chapter 11

Extending the study on the linear advection equation subject to stochastic velocity field and initial condition

In this chapter we extend the study on the linear advection equation with independent stochastic velocity and initial condition performed in [F.A. Dorini and M.C.C. Cunha. "On the linear advection equation subject to random velocity fields". In: Math. Comput. Simulat. 82 (2011), pp. 679–690]. By using both existing and novel results on the stochastic chain rule, we solve the random linear advection equation in the mean square sense. We provide a new expression for the probability density function of the solution stochastic process, which can be computed as accurate as wanted via Monte Carlo simulation, and which does not require the specific probability distribution of the integral of the velocity. This allows us to solve the non-Gaussian velocity case, which was not treated in the aforementioned contribution. Several numerical results illustrate the computations of the probability density function by using our approach. On the other hand, we derive a theoretical partial differential equation for the probability density function of the solution stochastic process. Finally, a shorter and easier derivation of the joint probability density function of the response process at two spatial points is obtained by applying conditional expectations appropriately.

11.1 Introduction

Advection equations arise in the modeling of a wide variety of physical processes that involve advective transport of substances or wave motions [54, 174]. An important model to describe the concentration of a chemical substance transported by a one-dimensional fluid that flows with a known velocity is formulated via the advection problem

$$\begin{cases} \frac{\partial}{\partial t}Q(x,t) + V(t)\frac{\partial}{\partial x}Q(x,t) = 0, \quad t > 0, \quad x \in \mathbb{R}, \\ Q(x,0) = Q_0(x), \quad x \in \mathbb{R}, \end{cases} \tag{11.1}$$

where $V(t)$ is the velocity and $Q_0(x)$ is the initial concentration of the substance at the spatial point. This partial differential equation has also been used to model the flux of a two-phase equal viscosity miscible fluid in a porous media.

When a real physical phenomenon is modeled via a differential equation, experiments are demanded to measure data and then to set the parameters of the differential equation. As a consequence, these parameters involve the uncertainty from error measurements. For example, in model (11.1) the total velocity, $V(t)$, is obtained from Darcy's law and it depends on the geology of the porous media. Therefore, the external velocity is defined by a given statistic. Moreover, the prediction of the initial state of the process is obtained by data acquired from a few number of exploratory wells by using geological methods [132]. This means that the interpretation of the data as random is notoriously more realistic. These facts motivate that the velocity $V(t)$ and the initial condition, $Q_0(x)$, in model (11.1) are better described via stochastic processes in lieu of deterministic functions. In this chapter, we will assume that these stochastic processes are defined in a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and that they are independent.

Several authors have studied problems related to model (11.1). Most of the approaches include methods by which one seeks the statistical moments (moment equations methods or average methods [54, 93, 152, 174]) and/or the probability density function (PDF methods [153, 166]) of the solution process. The main effort is usually concentrated on the derivation of appropriate differential equations for average quantities (moments) or the probability density function by using, in general, small perturbations with some closure approximation method [132]. Another approach is to solve appropriate equations for sets of realizations of random fields and to average computed functions, the so-called Monte Carlo approach, which has the advantage of applying to a very broad range of problems. The large volume of involved calculations, the slow

rate of convergence and the difficulty for generalizing the results may limit the importance of this method. To complete a fair overview of the state of the art with respect to the general goal of computing the probability density function of the solution process to random partial differential equations, it must be said that a powerful technique that has been extensively used is the Random Variable Transformation technique [130, 160]. Some relevant contributions in this regard include [64, 88, 89, 154], for example.

In this chapter we extend the study on the linear advection equation (11.1) performed in [58]. By using both extant and novel results on the stochastic chain rule, in the first part of Section 11.2 we solve the stochastic model (11.1) in the mean square sense. The second part of Section 11.2 is addressed to prove the uniqueness of the solution in the mean square sense. In Section 11.3, we provide a new expression for the probability density function of the solution process, which can be computed as accurate as wanted via Monte Carlo simulation, and which does not require the specific probability distribution of the integral of the velocity. This allows us to solve the non-Gaussian velocity case, which was not treated in [58]. Several numerical results illustrate the computations. A theoretical partial differential equation for the density function of the solution process is presented in Section 11.4. In Section 11.5, a shorter and easier derivation of the joint probability density function of the response process at two spatial points is obtained by applying conditional expectations appropriately. This section is completed by establishing a partial differential equation for the joint probability density function of the solution stochastic process to model (11.1). Finally, our conclusions are drawn in Section 11.6.

11.2 Mean square chain rule and application to solving the random linear advection equation

By using deterministic theory on partial differential equations [113], the sample-path solution to (11.1) is given by $Q(x,t,\omega) = Q_0(x - A(t,\omega))$, where $A(t,\omega) = \int_0^t V(\tau,\omega)\,\mathrm{d}\tau$ is a sample-path integral and ω is an outcome from our sample space Ω. This fact was already established in [58]. In this section, our purpose is to prove that $Q(x,t) = Q_0(x - A(t))$ is a mean square solution to (11.1) under certain conditions, where $A(t) = \int_0^t V(\tau)\,\mathrm{d}\tau$ is understood now as a mean square Riemann integral [160, 167]. Since $Q_0(x - A(t))$ is a composition of two stochastic processes, we need appropriate versions of the chain rule for mean square differentiation.

In the literature, the mean square differentiation of the composition of two stochastic processes has been studied when one of the two involved stochastic processes is a deterministic function:

Theorem 11.1 (Chain Rule Theorem, version 1) *[167, Th. 3.19] Let f be a deterministic C^1 function. Let $\{X(t) : t \in [a,b]\}$ be a stochastic process and $t \in [a,b]$ such that:*

(i) X is mean fourth differentiable at t.

(ii) X is path continuous on $[a,b]$.

(iii) There exist $r > 4$ and $\delta > 0$ such that $\sup_{s\in[-\delta,\delta]} \mathbb{E}[|f'(X(t+s))|^r] < \infty$.

Then $f(X(t))$ is mean square differentiable at t and $\frac{\mathrm{d}}{\mathrm{d}t} f(X(t)) = f'(X(t))X'(t)$.

Theorem 11.2 (Chain Rule Theorem, version 2) *[53, Th. 2.1] Let g be a deterministic differentiable function. Let $\{Y(t) : t \in [a,b]\}$ be a stochastic process such that $[a,b]$ contains the range of g. Suppose that Y is mean square C^1. Then $Y(g(t))$ is mean square differentiable on the whole domain of g and $\frac{\mathrm{d}}{\mathrm{d}t} Y(g(t)) = Y'(g(t))g'(t)$ for each t.*

In this chapter, we extend the chain rule theorem to the composition of two stochastic processes:

Theorem 11.3 (Chain Rule Theorem, general version) *Let $\{Y(t) : t \in [c,d]\}$ be a mean square differentiable stochastic process on $[c,d]$ with C^1 sample paths. Denote by Y' the mean square derivative of Y, and by $\dot{Y}$ the classical derivative of Y. Suppose that Y' and $\dot{Y}$ are indistinguishable stochastic process (i.e., $\mathbb{P}[Y'(t) = \dot{Y}(t),\ \forall t] = 1$). Let $\{X(t) : t \in [a,b]\}$ be a stochastic process with range in $[c,d]$, and $t \in [a,b]$ such that:*

(i) X is mean fourth differentiable at t.

(ii) X is path continuous on $[a,b]$.

(iii) There exist $r > 4$ and $\delta > 0$ such that $\sup_{s\in[-\delta,\delta]} \mathbb{E}[|Y'(X(t+s))|^r] < \infty$.

Then $Y(X(t))$ is mean square differentiable at t and $\frac{\mathrm{d}}{\mathrm{d}t} Y(X(t)) = Y'(X(t))X'(t)$.

Proof. The proof from [167, Th. 3.19] (i.e., Theorem 11.1) is applicable with $f = Y$. Then $(Y \circ X)'(t) = \dot{Y}(X(t))X'(t)$. But since Y' and $\dot{Y}$ are indistinguishable, $\dot{Y}(X(t)) = Y'(X(t))$ almost surely. Thus, $(Y \circ X)'(t) = Y'(X(t))X'(t)$, as wanted.

□

Remark 11.4 *In Theorem 11.3, if Y is a deterministic function f, then we obtain Theorem 11.1 again. Thus, Theorem 11.3 is an extension of Theorem 11.1.*

Remark 11.5 *The proof from [53, Th. 2.1] cannot be adapted to the composition of two stochastic processes. For example, formulas (2.7) and (2.9) therein cannot be generalized for a stochastic g.*

From the chain rule theorems for mean square differentiation, we can rigorously solve (11.1) in the mean square sense. We will state and prove two theorems. In the first one, we will consider a deterministic velocity field, so that Theorem 11.2 will be applicable. In the second theorem, both the velocity and the initial condition will be stochastic processes, and Theorem 11.3 will be applicable.

Theorem 11.6 *Let $V(t) = v(t)$ be a deterministic velocity function and $Q_0(x)$ be a stochastic initial condition. Assume that:*

(i) $v(t)$ is continuous on $[0, \infty)$.

(ii) $Q_0(x)$ is mean square $C^1(\mathbb{R})$.

Then $Q(x,t) = Q_0(x - A(t))$, with $A(t) = \int_0^t v(\tau)\mathrm{d}\tau$, is a mean square solution to (11.1).

Proof. Since $V(t) = v(t)$ is a deterministic function, the mean square Riemann integral $A(t) = \int_0^t v(\tau)\,\mathrm{d}\tau$ becomes an ordinary Riemann integral. By the classical Fundamental Theorem of Calculus, $A(t)$ is a $C^1([0,\infty))$ deterministic function and $A'(t) = v(t)$. Then, fixed $x \in \mathbb{R}$, $x - A(t)$ is a $C^1([0,\infty))$ deterministic function. By Theorem 11.2, $\frac{\partial}{\partial t}Q(x,t) = -Q_0'(x - A(t))A'(t) = -Q_0'(x - A(t))V(t)$. On the other hand, it is clear that $\frac{\partial}{\partial x}Q(x,t) = Q_0'(x - A(t))$. Hence, (11.1) is satisfied in the mean square setting.

□

Theorem 11.7 *Let $V(t)$ be a stochastic velocity and $Q_0(x)$ be a stochastic initial condition. Suppose that:*

(i) $Q_0(x)$ is mean square differentiable on $\mathbb{R}$, with sample paths in $C^1(\mathbb{R})$, and with mean square derivative and classical derivative being indistinguishable stochastic processes.

(ii) $V(t)$ is mean fourth continuous on $[0,\infty)$.

(iii) We have

$$\sup_{s\in[-\delta,\delta]} \mathbb{E}[|Q_0'(x-A(t+s))|^r] < \infty, \text{ for some } r>4 \text{ and } \delta>0$$

and

$$\sup_{h\in[-\eta,\eta]} \mathbb{E}[|Q_0'(x+h-A(t))|^q] < \infty, \text{ for some } q>4 \text{ and } \eta>0,$$

for each $x\in\mathbb{R}$ and $t>0$.

Then $Q(x,t)=Q_0(x-A(t))$, with $A(t)=\int_0^t V(\tau)\mathrm{d}\tau$, is a mean square solution to (11.1).

Proof. Let us see that $\frac{\partial}{\partial x}Q(x,t)=Q_0'(x-A(t))$. Fix $t>0$. Let the process $X(x)=x-A(t)$. Clearly, $X(x)$ is mean fourth differentiable with continuous sample paths, so conditions (i) and (ii) from Theorem 11.3 hold. On the other hand,

$$\sup_{h\in[-\eta,\eta]} \mathbb{E}[|Q_0'(X(x+h))|^q] < \infty, \text{ for some } q>4 \text{ and } \eta>0,$$

for each $x\in\mathbb{R}$, so condition (iii) from Theorem 11.3 holds. Then, by Theorem 11.3, there exists $\frac{\partial}{\partial x}Q(x,t)=\frac{\partial}{\partial x}(Q_0(X(x)))=Q_0'(X(x))X'(x)=Q_0'(x-A(t))$, as wanted.
Since $V(t)$ is mean fourth continuous, by [167, Prop. 3.18] we derive that $A(t)$ is mean fourth C^1 with $A'(t)=V(t)$ (Fundamental Theorem of mean fourth calculus). Let us see that $\frac{\partial}{\partial t}Q(x,t)=-Q_0'(x-A(t))V(t)$. Fix $x\in\mathbb{R}$. Let $X(t)=x-A(t)$. We have that $X(t)$ is mean fourth C^1 on $(0,\infty)$, with $X'(t)=-A'(t)=-V(t)$. By [45, Lemma 4], $X(t)$ has continuous sample paths on $[0,\infty)$. Thus, conditions (i) and (ii) from Theorem 11.3 are satisfied. Finally,

$$\sup_{s\in[-\delta,\delta]} \mathbb{E}[|Q_0'(X(t+s))|^r] < \infty, \text{ for some } r>4 \text{ and } \delta>0,$$

for each $t > 0$, which is condition (iii) from Theorem 11.3. Hence, by Theorem 11.3, $\frac{\partial}{\partial t}Q(x,t) = \frac{\partial}{\partial t}Q_0(X(t)) = Q_0'(X(t))X'(t) = -Q_0'(x - A(t))V(t)$. In conclusion, equation (11.1) holds.

□

Example 11.8 Let $V(t)$ be a standard Brownian motion on $[0,\infty)$ (then $V(t) \sim \text{Normal}(0,t)$) [160, Section 3.3.4], $Q_0(x)$ be the deterministic smooth function $g(x) = \alpha e^{-\beta x^2}$, for certain $\alpha, \beta > 0$. These random inputs have been already used in [58, Example 3.2]. By [167, Lemma 3.11], $V(t)$ is mean fourth continuous on $[0,\infty)$ if and only if $\mathbb{E}[V(t_1)V(t_2)V(t_3)V(t_4)]$ is continuous at each $(t,t,t,t) \in [0,\infty)^4$. This is clear by property [160, p. 28, formula (2.101)] and the covariance function of Brownian motion, given by $\mathbb{E}[V(t_i)V(t_j)] = \min\{t_i, t_j\}$. Thus, conditions (i) and (ii) from Theorem 11.7 hold. Finally, for condition (iii) from Theorem 11.7, we compute $g'(x) = -2\alpha\beta x e^{-\beta x^2}$, which is bounded in absolute value by a number $M_{\alpha,\beta} > 0$ uniformly on $x \in \mathbb{R}$. By Theorem 11.7, $Q(x,t) = Q_0(x - A(t))$ is a rigorous mean square solution to (11.1).

Example 11.9 Consider $V(t) = \arctan(e^{Bt} - 1)$, where $B \sim \text{Binomial}(3, 0.1)$, and $Q_0(x) = \sin(Cx)$, where $C \sim \text{Poisson}(3)$. The random variables B and C are assumed to be independent. The initial condition Q_0 has $C^1(\mathbb{R})$ sample paths, because the sine function is smooth. On the other hand, the mean square differentiability property of Q_0 follows from Theorem 11.1, with $f = \sin$ and $X(x) = Cx$. Moreover, the mean square and the classical derivatives of Q_0 coincide, because of the chain rule formula. This shows condition (i) from Theorem 11.7. For condition (ii) from Theorem 11.7, we use the facts that $V(t+h,\omega) - V(t,\omega) \to 0$ as $h \to 0$ almost surely and that $|V(t+h,\omega) - V(t,\omega)| \leq |V(t+h,\omega)| + |V(t,\omega)| \leq \pi$ almost surely, so the Dominated Convergence Theorem implies that $\|V(t+h,\omega) - V(t,\omega)\|_{\mathrm{L}^4(\Omega)} \to 0$ as $h \to 0$, as wanted. Another possible approach to check that condition (ii) holds true could consist in using [45, Lemma 5], with $f(y) = \arctan(e^y - 1)$ and $Y(t) = Bt$ in the notation therein. Finally, for condition (iii), $\mathbb{E}[|Q_0'(x)|^p] \leq \mathbb{E}[C^p] < \infty$. Thus, Theorem 11.7 implies that $Q(x,t) = Q_0(x - A(t))$ is a rigorous mean square solution to (11.1).

Example 11.10 Let $V(t)$ be a standard Brownian motion on $[0,\infty)$ (then $V(t) \sim \text{Normal}(0,t)$) [160, Section 3.3.4], $Q_0(x) = Be^x$, $B \sim \text{Exponential}(2)$. Condition (i) is clear. Condition (ii) also holds, see Example 11.8. Finally, condition (iii) holds: indeed, it is well known that $A(t) \sim \text{Normal}(0, \frac{t^3}{3})$ (see

[160, Example 4.12] taking diffusion constant $2D \equiv 1$), then by using the finiteness of the moment generating function of a normal distribution, and by Cauchy-Schwarz inequality, we deduce that

$$\mathbb{E}[|Q_0'(x - A(t+s))|^r] = \mathbb{E}[B^r \mathrm{e}^{r(x-A(t+s))}] \leq \mathbb{E}[B^{2r}]^{\frac{1}{2}} \mathbb{E}[\mathrm{e}^{2r(x-A(t+s))}]^{\frac{1}{2}} < \infty,$$

$$\mathbb{E}[|Q_0'(x + h - A(t))|^q] = \mathbb{E}[B^q \mathrm{e}^{q(x+h-A(t))}] \leq \mathbb{E}[B^{2q}]^{\frac{1}{2}} \mathbb{E}[\mathrm{e}^{2q(x+h-A(t))}]^{\frac{1}{2}} < \infty.$$

A similar reasoning has been used in [167, Example 3.20]. By Theorem 11.7, the process $Q(x,t) = Q_0(x - A(t))$ is a rigorous mean square solution to (11.1).

To finish this section, we deal with uniqueness, in the mean square sense, of the solution stochastic process to system (11.1). Suppose that $Q_1(x,t)$ and $Q_2(x,t)$ are two mean square solutions to (11.1) that belong to $C^1(\mathbb{R} \times [0,\infty))$ in the mean square sense. Let $\bar{Q}(x,t) = Q_1(x,t) - Q_2(x,t)$. Then

$$\begin{cases} \frac{\partial}{\partial t}\bar{Q}(x,t) + V(t)\frac{\partial}{\partial x}\bar{Q}(x,t) = 0, \quad t > 0, \quad x \in \mathbb{R}, \\ \bar{Q}(x,0) = 0, \quad x \in \mathbb{R}. \end{cases} \tag{11.2}$$

The goal is to demonstrate that $\bar{Q}(x,t) = 0$ almost surely, for each $x \in \mathbb{R}$ and $t \geq 0$ (that is, $\bar{Q}$ is an equivalent stochastic process to 0).

If we prove that

$$\frac{\partial}{\partial t}\left(\bar{Q}(x + A(t), t)\right) = \frac{\partial \bar{Q}}{\partial t}(x + A(t), t) + V(t)\frac{\partial \bar{Q}}{\partial x}(x + A(t), t), \tag{11.3}$$

then by (11.2),

$$\frac{\partial}{\partial t}\left(\bar{Q}(x + A(t), t)\right) = 0.$$

If we multiply by $\bar{Q}(x + A(t), t)$ and use the product rule for mean square differentiation (variation of [167, Lemma 3.14]), we derive that

$$\frac{\partial}{\partial t}\left(\bar{Q}(x + A(t), t)^2\right) = 2\bar{Q}(x + A(t), t)\frac{\partial}{\partial t}\left(\bar{Q}(x + A(t), t)\right) = 0,$$

where the first derivative is understood in the $\mathrm{L}^1(\Omega)$ sense. By [160, p. 97],

$$\frac{\partial}{\partial t}\mathbb{E}\left[\bar{Q}(x + A(t), t)^2\right] = \mathbb{E}\left[\frac{\partial}{\partial t}\left(\bar{Q}(x + A(t), t)^2\right)\right] = 0.$$

This implies that $t \mapsto \mathbb{E}\left[\bar{Q}(x + A(t), t)^2\right]$ is constant. From the initial condition in (11.2), $\mathbb{E}\left[\bar{Q}(x + A(t), t)^2\right] = \mathbb{E}\left[\bar{Q}(x, 0)^2\right] = 0$, which entails $\bar{Q}(x +$

$A(t), t) = 0$ almost surely. As x and t are arbitrary, we conclude that $\bar{Q}(x,t) = 0$ almost surely, for each $x \in \mathbb{R}$ and $t \geq 0$, as desired.

Thus, the key result to be proved is (11.3). It corresponds to a mean square chain rule but with 2 variables involved. By definition of mean square differentiation, (11.3) is equivalent to

$$\lim_{h\to 0} \left\| \frac{\bar{Q}(x+A(t+h),t+h) - \bar{Q}(x+A(t),t)}{h} - \frac{\partial \bar{Q}}{\partial t}(x+A(t),t) - V(t)\frac{\partial \bar{Q}}{\partial x}(x+A(t),t) \right\|_{\mathrm{L}^2(\Omega)} = 0, \tag{11.4}$$

for each $x \in \mathbb{R}$ and $t \geq 0$. Fix x and t. We start by bounding

$$\begin{aligned}
&\left\| \frac{\bar{Q}(x+A(t+h),t+h) - \bar{Q}(x+A(t),t)}{h} - \frac{\partial \bar{Q}}{\partial t}(x+A(t),t) - V(t)\frac{\partial \bar{Q}}{\partial x}(x+A(t),t) \right\|_{\mathrm{L}^2(\Omega)} \\
\leq &\left\| \frac{\bar{Q}(x+A(t+h),t+h) - \bar{Q}(x+A(t+h),t)}{h} - \frac{\partial \bar{Q}}{\partial t}(x+A(t),t) \right\|_{\mathrm{L}^2(\Omega)} \\
+ &\left\| \frac{\bar{Q}(x+A(t+h),t) - \bar{Q}(x+A(t),t)}{h} - V(t)\frac{\partial \bar{Q}}{\partial x}(x+A(t),t) \right\|_{\mathrm{L}^2(\Omega)},
\end{aligned}$$

whence (11.4) is deduced from

$$\lim_{h\to 0} \left\| \frac{\bar{Q}(x+A(t+h),t+h) - \bar{Q}(x+A(t+h),t)}{h} - \frac{\partial \bar{Q}}{\partial t}(x+A(t),t) \right\|_{\mathrm{L}^2(\Omega)} = 0. \tag{11.5}$$

$$\lim_{h\to 0} \left\| \frac{\bar{Q}(x+A(t+h),t) - \bar{Q}(x+A(t),t)}{h} - V(t)\frac{\partial \bar{Q}}{\partial x}(x+A(t),t) \right\|_{\mathrm{L}^2(\Omega)} = 0. \tag{11.6}$$

To deal with (11.5), we use the fundamental theorem of mean square calculus [160, p. 104(6)]:

$$\begin{aligned}
&\frac{\bar{Q}(x+A(t+h),t+h) - \bar{Q}(x+A(t+h),t)}{h} = \frac{1}{h}\int_t^{t+h} \frac{\partial \bar{Q}}{\partial t}(x+A(t+h),s)\,\mathrm{d}s \\
= &\frac{1}{h}\int_t^{t+h} \left(\frac{\partial \bar{Q}}{\partial t}(x+A(t+h),s) - \frac{\partial \bar{Q}}{\partial t}(x+A(t),s) \right) \mathrm{d}s \\
+ &\frac{1}{h}\int_t^{t+h} \frac{\partial \bar{Q}}{\partial t}(x+A(t),s)\,\mathrm{d}s
\end{aligned}$$

(here we are using mean square Riemann integrals). Since the map

$$s \mapsto \frac{\partial \bar{Q}}{\partial t}(x+A(t),s)$$

is mean square continuous, by [160, p. 103(5)]

$$\lim_{h\to 0} \frac{1}{h}\int_t^{t+h} \frac{\partial \bar{Q}}{\partial t}(x+A(t),s)\,\mathrm{d}s = \frac{\partial \bar{Q}}{\partial t}(x+A(t),t) \tag{11.7}$$

in $\mathrm{L}^2(\Omega)$. On the other hand, we need

$$\lim_{h\to 0}\frac{1}{h}\int_t^{t+h}\left(\frac{\partial\bar{Q}}{\partial t}(x+A(t+h),s)-\frac{\partial\bar{Q}}{\partial t}(x+A(t),s)\right)\mathrm{d}s=0 \tag{11.8}$$

in $\mathrm{L}^2(\Omega)$. This would be deduced if we are able to prove that

$$(s_1,s_2)\mapsto\frac{\partial\bar{Q}}{\partial t}(x+A(s_2),s_1)\text{ is mean square continuous.} \tag{11.9}$$

Indeed, from (11.9), one has that $(s_1,s_2)\mapsto\frac{\partial\bar{Q}}{\partial t}(x+A(s_2),s_1)$ is mean square uniformly continuous on $[t-1,t+1]^2\cap[0,\infty)^2$: given any $\epsilon>0$, there is a $0<\mu<1$ such that, if $|s_3-s_1|<\mu$ and $|s_4-s_2|<\mu$ with $s_1,s_2,s_3,s_4\in[t-1,t+1]^2\cap[0,\infty)^2$, then

$$\left\|\frac{\partial\bar{Q}}{\partial t}(x+A(s_2),s_1)-\frac{\partial\bar{Q}}{\partial t}(x+A(s_4),s_3)\right\|_{\mathrm{L}^2(\Omega)}<\epsilon. \tag{11.10}$$

If $|h|<\mu$, then from [167, Prop. 3.17] in the setting of mean square calculus (or [160, p. 102(3)]) and (11.10),

$$\begin{aligned}&\frac{1}{h}\left\|\int_t^{t+h}\left(\frac{\partial\bar{Q}}{\partial t}(x+A(t+h),s)-\frac{\partial\bar{Q}}{\partial t}(x+A(t),s)\right)\mathrm{d}s\right\|_{\mathrm{L}^2(\Omega)}\\&\le\frac{1}{h}\int_t^{t+h}\left\|\frac{\partial\bar{Q}}{\partial t}(x+A(t+h),s)-\frac{\partial\bar{Q}}{\partial t}(x+A(t),s)\right\|_{\mathrm{L}^2(\Omega)}\mathrm{d}s<\epsilon,\end{aligned}$$

which shows (11.8). This together with the previous expression (11.7) gives (11.5).

Therefore, only (11.9) and (11.6) remain to be proved. We do so in the following two theorems by assuming certain hypotheses.

Theorem 11.11 *Let $V(t)=v(t)$ be a deterministic velocity function and $Q_0(x)$ be any stochastic initial condition. Suppose that $v(t)$ is continuous on $[0,\infty)$. Then there is at most one stochastic process $Q(x,t)$ (up to equivalence) that is mean square $C^1(\mathbb{R}\times[0,\infty))$ and solves (11.1) in the mean square sense.*

Proof. Suppose that $Q_1(x,t)$ and $Q_2(x,t)$ are two mean square solutions to (11.1) that belong to $C^1(\mathbb{R}\times[0,\infty))$ in the mean square sense, and consider $\bar{Q}(x,t)=Q_1(x,t)-Q_2(x,t)$. We need to prove both (11.9) and (11.6).
Since $V(t)=v(t)$ is a deterministic function, the mean square Riemann integral $A(t)=\int_0^t v(\tau)\,\mathrm{d}\tau$ becomes an ordinary Riemann integral. By the classical

Fundamental Theorem of Calculus, $A(t)$ is a $C^1([0,\infty))$ deterministic function and $A'(t) = v(t)$.
To demonstrate (11.9), notice that if $s_{1,n} \to s_1$ and $s_{2,m} \to s_2$ as $n, m \to \infty$, then $A(s_{2,m}) \to A(s_2)$ by deterministic continuity, therefore

$$\frac{\partial \bar{Q}}{\partial t}(x + A(s_{2,m}), s_{1,n}) \overset{n,m\to\infty}{\longrightarrow} \frac{\partial \bar{Q}}{\partial t}(x + A(s_2), s_1)$$

in $\mathrm{L}^2(\Omega)$, by mean square continuity of $\frac{\partial \bar{Q}}{\partial t}$ on $\mathbb{R} \times [0,\infty)$.
To prove (11.6), let $Y(y) = \bar{Q}(y,t)$ and $g(s) = x + A(t+s)$ (recall that x and t are fixed). Notice that (11.6) is equivalent to $(Y \circ g)'(0) = Y'(g(0))g'(0)$, therefore the conditions of Theorem 11.2 must be met. Since A is differentiable, g is also differentiable. As $\bar{Q}(x,t)$ is mean square $C^1(\mathbb{R} \times [0,\infty)$, the process Y is mean square $C^1(\mathbb{R})$. Theorem 11.2 applies to justify $(Y \circ g)'(0) = Y'(g(0))g'(0)$, as wanted.

□

Theorem 11.12 *Let $V(t)$ be a mean fourth continuous process on $[0,\infty)$ and $Q_0(x)$ be any stochastic initial condition. Consider the following three properties:*

1. *$Q(x,t)$ has mean square partial derivatives, it has sample paths in $C^1(\mathbb{R}\times[0,\infty))$, and its mean square partial derivatives and classical partial derivatives are indistinguishable stochastic processes on $\mathbb{R} \times [0,\infty)$.*

2. *It holds*

$$\sup_{s\in[-\delta,\delta]} \mathbb{E}\left[\left|\frac{\partial Q}{\partial x}(x + A(t+s), t)\right|^r\right] < \infty, \text{ for some } r > 4 \text{ and } \delta > 0,$$

   for each $x \in \mathbb{R}$ and $t \geq 0$.

3. *It holds*

$$\sup_{r_1,r_2\in[-\eta,\eta]} \mathbb{E}\left[\left|\frac{\partial Q}{\partial t}(x + A(s_2 + r_2), s_1 + r_1)\right|^q\right] < \infty, \text{ for some } q > 2,\ \eta > 0,$$

   for each $x \in \mathbb{R}$ and $s_1, s_2 \geq 0$.

Then there is at most one stochastic process $Q(x,t)$ (up to equivalence) that satisfies these three properties and solves (11.1) in the mean square sense.

Proof. Suppose that $Q_1(x,t)$ and $Q_2(x,t)$ are two mean square solutions to (11.1) that satisfy the three properties, and consider $\bar{Q}(x,t) = Q_1(x,t) -$

$Q_2(x,t)$. By [167, Prop. 2.3] (the so-called c_s-inequality), $\bar{Q}(x,t)$ also satisfies the three properties. We need to prove both (11.9) and (11.6).
Since $V(t)$ is mean fourth continuous, by [167, Prop. 3.18] we derive that $A(t)$ is mean fourth C^1 with $A'(t) = V(t)$ (Fundamental Theorem of mean fourth calculus). By [45, Lemma 4], $A(t)$ has continuous sample paths on $[0,\infty)$. Since $\bar{Q}(x,t)$ possesses sample paths in $C^1(\mathbb{R}\times[0,\infty))$, the map

$$(s_1,s_2) \mapsto \frac{\partial \bar{Q}}{\partial t}(x + A(s_2), s_1)$$

from (11.9) is almost surely continuous. Consider two sequences $s_{1,n} \to s_1$ and $s_{2,m} \to s_2$ as $n,m \to \infty$. We have

$$\frac{\partial \bar{Q}}{\partial t}(x + A(s_{2,m}), s_{1,n}) \overset{n,m\to\infty}{\longrightarrow} \frac{\partial \bar{Q}}{\partial t}(x + A(s_2), s_1) \tag{11.11}$$

almost surely, therefore in probability. By the third property and [167, Th. 2.4], (11.11) holds in $\mathrm{L}^2(\Omega)$, and we are done with (11.9).
To demonstrate (11.6), let $Y(y) = \bar{Q}(y,t)$ and $X(z) = x + A(t+z)$ (recall that x and t are fixed). Notice that (11.6) is equivalent to $(Y\circ X)'(0) = Y'(X(0))X'(0)$, therefore the conditions of Theorem 11.3 must be met. We know that $A(t)$ is mean fourth differentiable with continuous sample paths, then so is $X(t)$, and conditions (i) and (ii) from Theorem 11.3 hold. Finally, condition (iii) therein is a direct consequence of our second hypothesis. Theorem 11.3 applies and $(Y\circ X)'(0) = Y'(X(0))X'(0)$ holds, as wanted.

□

We state and prove a last uniqueness theorem with different techniques than those previously used. It is based on an energy method somewhat. We apply it for processes solving (11.1) that are periodic on x. For example, if $Q(x,t) = Dt\sin(2\pi x) + E$, where D and E are random variables, then $Q(x,t)$ is 1-periodic on x, since $Q(x+1,t) = Q(x,t)$ almost surely for each $x\in\mathbb{R}$ and $t \geq 0$. Notice that if $Q_0(x)$ is p-periodic, then $Q(x,t) = Q_0(x - A(t))$ is p-periodic on x too.

Theorem 11.13 *Let $V(t)$ be any bounded stochastic velocity ($V(t)\in\mathrm{L}^\infty(\Omega)$, for each $t\geq 0$) and let $Q_0(x)$ be any initial condition process. Let $p>0$. Then there is at most one mean square solution to (11.1) (up to equivalence) that is mean square $C^1(\mathbb{R}\times[0,\infty))$ and p-periodic on x.*

Proof. Suppose that $Q_1(x,t)$ and $Q_2(x,t)$ are two mean square solutions to (11.1) being mean square $C^1(\mathbb{R}\times[0,\infty))$ and p-periodic on x, and consider the p-periodic solution $\bar{Q}(x,t) = Q_1(x,t) - Q_2(x,t)$ to (11.2).

From (11.2) and by using the product rule for mean square differentiation (variation of [167, Lemma 3.14]), we obtain the random partial differential equation

$$\frac{\partial}{\partial t}\left(\bar{Q}(x,t)^2\right) + V(t)\frac{\partial}{\partial x}\left(\bar{Q}(x,t)^2\right) = 0,$$

where the partial derivatives are understood in the $\mathrm{L}^1(\Omega)$ sense. Now we apply Riemann integrals in the $\mathrm{L}^1(\Omega)$ sense. We use

$$\int_{x_1}^{x_1+p} V(t)\frac{\partial}{\partial x}\left(\bar{Q}(x,t)^2\right)\,\mathrm{d}x = V(t)\int_{x_1}^{x_1+p}\frac{\partial}{\partial x}\left(\bar{Q}(x,t)^2\right)\,\mathrm{d}x,$$

as a consequence of the boundedness of $V(t)$, and

$$\int_{x_1}^{x_1+p}\frac{\partial}{\partial x}\left(\bar{Q}(x,t)^2\right)\,\mathrm{d}x = \bar{Q}(x_1+p,t)^2 - \bar{Q}(x_1,t)^2 = 0$$

(Barrow's rule in $\mathrm{L}^1(\Omega)$ is justified by $\mathrm{L}^1(\Omega)$-continuity of $\frac{\partial}{\partial x}\left(\bar{Q}(x,t)^2\right)$, see [160, p. 104(6)]), therefore

$$\int_{x_1}^{x_1+p} V(t)\frac{\partial}{\partial x}\left(\bar{Q}(x,t)^2\right)\,\mathrm{d}x = 0.$$

We also use

$$\int_{x_1}^{x_1+p}\frac{\partial}{\partial t}\left(\bar{Q}(x,t)^2\right)\,\mathrm{d}x = \frac{\partial}{\partial t}\int_{x_1}^{x_1+p}\bar{Q}(x,t)^2\,\mathrm{d}x,$$

see [45, Prop. 6]. Hence,

$$\frac{\partial}{\partial t}\int_{x_1}^{x_1+p}\bar{Q}(x,t)^2\,\mathrm{d}x = 0.$$

By applying expectations,

$$\frac{\partial}{\partial t}\int_{x_1}^{x_1+p}\mathbb{E}\left[\bar{Q}(x,t)^2\right]\,\mathrm{d}x = 0.$$

Then the real map $t \mapsto \int_{x_1}^{x_1+p}\mathbb{E}\left[\bar{Q}(x,t)^2\right]\,\mathrm{d}x$ is constant, therefore

$$\int_{x_1}^{x_1+p}\mathbb{E}\left[\bar{Q}(x,t)^2\right]\,\mathrm{d}x = \int_{x_1}^{x_1+p}\mathbb{E}\left[\bar{Q}(x,0)^2\right]\,\mathrm{d}x = 0.$$

This implies $\mathbb{E}\left[\bar{Q}(x,t)^2\right] = 0$, for $x \in \mathbb{R}$ and $t \geq 0$. Thereby, $\bar{Q}(x,t) = 0$ almost surely, for each $x \in \mathbb{R}$ and $t \geq 0$, as wanted.

□

11.3 Density function of the solution process

In this section, the goal is to compute or approximate the probability density function $f_Q(q;x,t)$ of the solution process $Q(x,t) = Q_0(x-A(t))$, where $A(t) = \int_0^t V(\tau)\,\mathrm{d}\tau$ can be understood as a sample-path integral or as a mean square Riemann integral. The knowledge of the probability density function of $Q(x,t)$ is important to perform uncertainty quantification. From it, the main statistics of $Q(x,t)$ can be determined, say the expectation and variance:

$$\mathbb{E}[Q(x,t)] = \int_{\mathbb{R}} q f_Q(q;x,t)\,\mathrm{d}q,\ \mathbb{V}[Q(x,t)] = \int_{\mathbb{R}} q^2 f_Q(q;x,t)\,\mathrm{d}q - \left(\mathbb{E}[Q(x,t)]\right)^2. \tag{11.12}$$

In [58, Prop. 2.1], the authors justify $f_Q(q;x,t) = \int_{\mathbb{R}} f_{A(t)}(x-x_0) f_{Q_0}(q;x_0)\,\mathrm{d}x_0$, by using the law of total probability. As the authors commented in [58, Remark 2.1], this expression requires the knowledge of the probability density function of $A(t)$, but the density function of the integral of a stochastic process $V(t)$ is generally not known, unless $V(t)$ is Gaussian [160, Th. 4.6.4]. Thus, all the examples presented in [58] were restricted to Gaussian velocities.

We will show an alternative application of the law of total probability so that we will be able to compute $f_Q(q;x,t)$ as long as exact realizations of $A(t)$ can be generated, independent of $V(t)$ being Gaussian or not. Instead of conditioning to $X_0(x,t) = x - A(t)$, we condition directly to $A(t)$. If we denote by $\mathbb{P}_{A(t)} = \mathbb{P} \circ (A(t))^{-1}$ the probability law of $A(t)$ (which may be absolutely continuous, discrete or singular), then

$$\begin{aligned}
F_Q(q;x,t) &= \mathbb{P}(Q(x,t) \le q) = \mathbb{P}(Q_0(x - A(t)) \le q) \\
&= \int_{\mathbb{R}} \mathbb{P}(Q_0(x - A(t)) \le q | A(t) = a)\, \mathbb{P}_{A(t)}(\mathrm{d}a) \\
&= \int_{\mathbb{R}} \mathbb{P}(Q_0(x - a) \le q | A(t) = a)\, \mathbb{P}_{A(t)}(\mathrm{d}a) \\
&= \int_{\mathbb{R}} \mathbb{P}(Q_0(x - a) \le q)\, \mathbb{P}_{A(t)}(\mathrm{d}a) \\
&= \int_{\mathbb{R}} F_{Q_0}(q; x - a)\, \mathbb{P}_{A(t)}(\mathrm{d}a) \\
&= \mathbb{E}[F_{Q_0}(q; x - A(t))],
\end{aligned} \tag{11.13}$$

where the independence between the velocity and the initial condition processes, V and Q_0, has been used. Hence,

$$f_Q(q;x,t) = \int_{\mathbb{R}} f_{Q_0}(q; x - a)\, \mathbb{P}_{A(t)}(\mathrm{d}a) = \mathbb{E}[f_{Q_0}(q; x - A(t))]. \tag{11.14}$$

In the case that $A(t)$ is absolutely continuous, that is, $\mathbb{P}_{A(t)}(\mathrm{d}a) = f_{A(t)}(a)\,\mathrm{d}a$, and by using the commutativity property of the convolution, we obtain again the formula from [58, Prop. 2.1]. Therefore, the previous development generalizes the one exhibited in [58].

If we can compute exact realizations of $A(t)$, say $A^{(1)}(t), \ldots, A^{(M)}(t)$, then the law of large numbers implies that

$$f_Q(q; x, t) = \mathbb{E}[f_{Q_0}(q; x - A(t))] \approx \frac{1}{M} \sum_{i=1}^{M} f_{Q_0}(q; x - A^{(i)}(t)). \tag{11.15}$$

Notice that the realizations from $A(t)$ can be computed from realizations of $V(t)$, as $A(t, \omega) = \int_0^t V(\tau, \omega)\,\mathrm{d}\tau$.

Example 11.14 Consider as initial condition process $Q_0(x) = Gx^2$, where $G \sim \text{Gamma}(3, 1)$. We take the following non-Gaussian velocity field: $V(t) = \cos(U\mathrm{e}^{Bt} + E)$, where $U \sim \text{Uniform}(0, 1)$, $B \sim \text{Binomial}(3, 0.3)$ and $E \sim \text{Exponential}(3)$. All the involved random variables are assumed to be independent. The integral defining $A(t)$ cannot be explicitly computed, and its probability distribution is not exactly known. We use (11.15) to approximate $f_Q(q; x, t)$. To illustrate our approach, we work at the points $x = -1, 0, 1, 3$ and $t = 0.4$. In Figure 11.1, we plot the graph of $f_Q(q; -1, 0.4)$, $f_Q(q; 0, 0.4)$, $f_Q(q; 1, 0.4)$ and $f_Q(q; 3, 0.4)$, respectively. From the probability density function, one can easily compute the main statistics of $Q(x, t)$ by utilizing (11.12).

Example 11.15 We take $Q_0(x) = Gx^2$, where $G \sim \text{Gamma}(3, 1)$ as in the previous example. Consider $V(t) = \cos(\mathrm{e}^{Bt} + 1)$, where $B \sim \text{Binomial}(3, 0.3)$. In this case, the velocity is a discrete random variable at each time instant. By using (11.15), we approximate $f_Q(q; x, t)$ at $x = -1, 0, 1, 3$ and $t = 0.4$. Figure 11.2 shows the graph of $f_Q(q; -1, 0.4)$, $f_Q(q; 0, 0.4)$, $f_Q(q; 1, 0.4)$ and $f_Q(q; 3, 0.4)$, respectively. By using (11.12), one can easily calculate the main statistics of $Q(x, t)$.

It may be possible that one cannot compute exact realizations of $A(t)$. For example, if $V(t) = \sin(B_t)$, where B_t is a standard Brownian motion on $[0, \infty)$, then we cannot compute exact realizations of $V(t)$, because the sample paths of B_t cannot be exactly computed. In such a case, a computational method to approximate the realizations of $V(t)$ as accurate as wanted is by truncating Karhunen-Loève expansions [119, Ch. 5], for example.

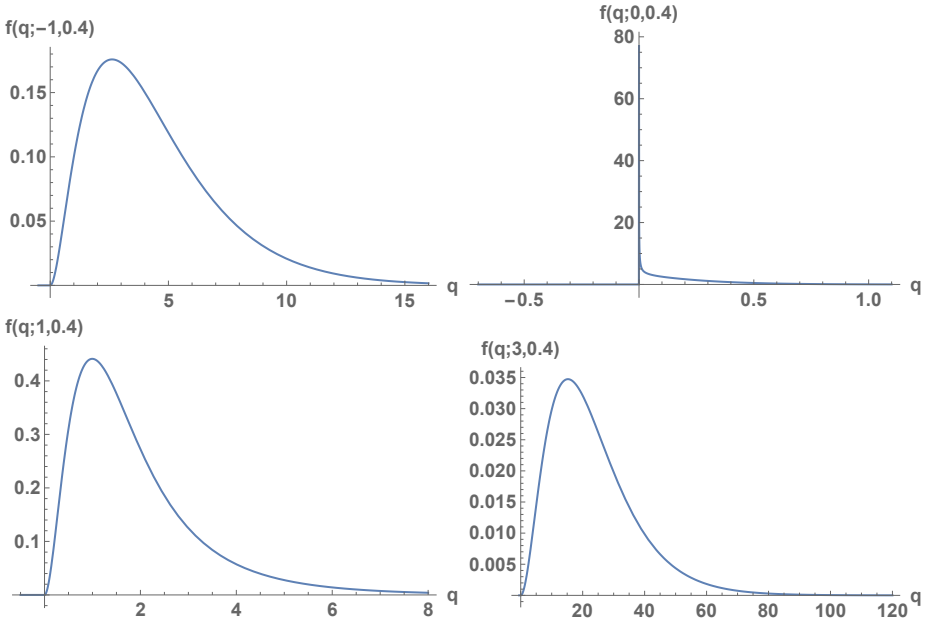

Figure 11.1: Graph of $f_Q(q;-1,0.4)$, $f_Q(q;0,0.4)$, $f_Q(q;1,0.4)$ and $f_Q(q;3,0.4)$ in Example 11.14.

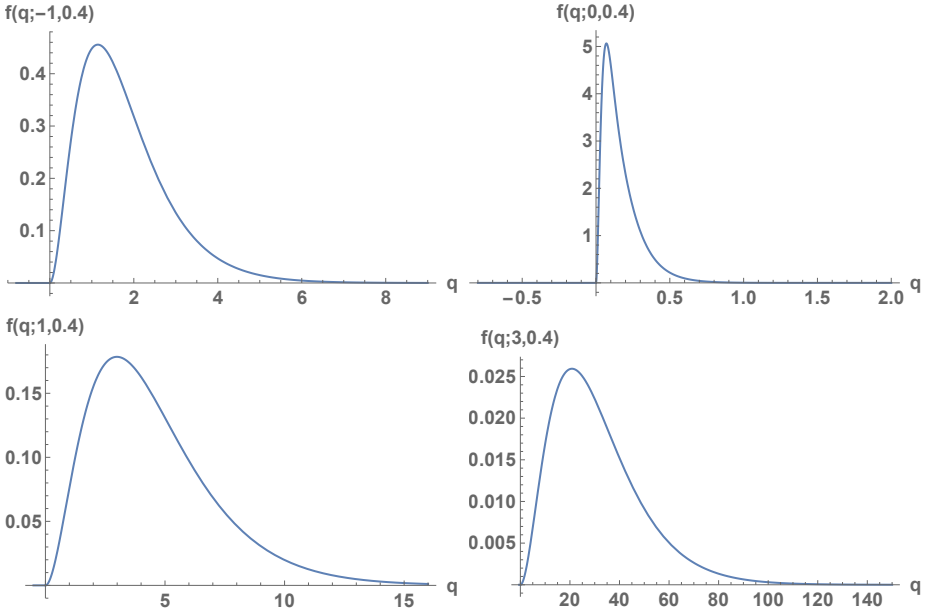

Figure 11.2: Graph of $f_Q(q;-1,0.4)$, $f_Q(q;0,0.4)$, $f_Q(q;1,0.4)$ and $f_Q(q;3,0.4)$ in Example 11.15.

Example 11.16 We consider $Q_0(x) = Gx^2$, where $G \sim \text{Gamma}(3,1)$ as in the previous two examples. We pick $V(t) = \sin(B_t)$, where B_t is a standard Brownian motion on $[0,\infty)$. This velocity is given by a non-Gaussian stochastic process. Our goal is to approximate $f_Q(q;0,0.4)$, $f_Q(q;0.5,0.4)$ and $f_Q(q;1,0.4)$ with (11.15). From the Karhunen-Loève expansion of Brownian motion on $[0,1]$, see [119, Exercise 5.12], we approximate

$$V(t) \approx \sin\left(\sum_{j=1}^{J} \frac{\sqrt{2}}{(j-1/2)\pi} \sin((j-1/2)\pi t)\xi_j\right), \quad t \in [0,1],$$

where $\xi_1, \ldots, \xi_J$ are independent Normal$(0,1)$ random variables. We use this expression to compute approximate realizations of $V(t)$, and therefore of $A(t)$. Figure 11.3 depicts the graph of $f_Q(q;0,0.4)$, $f_Q(q;0.5,0.4)$ and $f_Q(q;1,0.4)$. From the density function, one may compute the expectation and variance of $Q(x,t)$ by using (11.12).

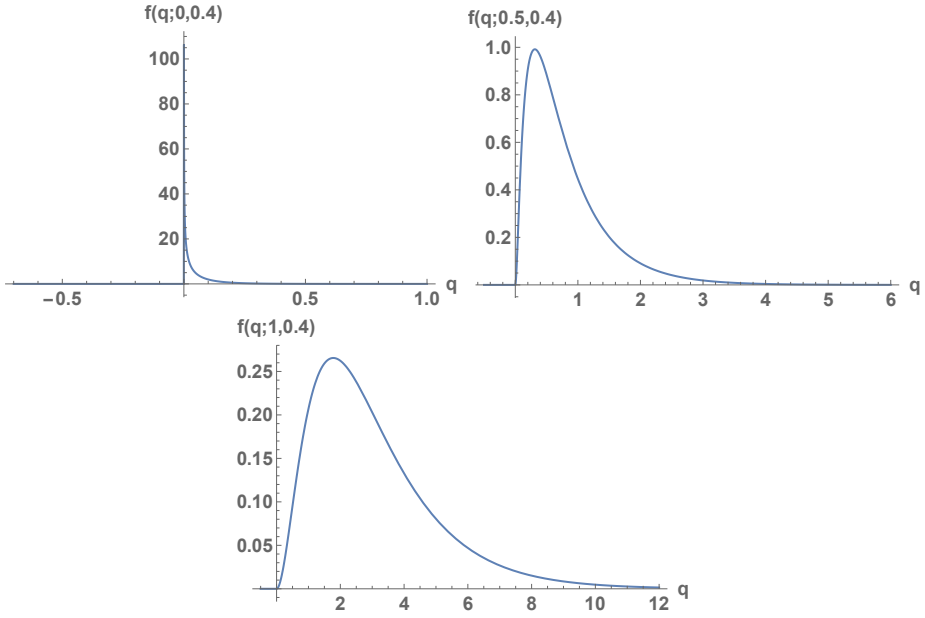

Figure 11.3: Graph of $f_Q(q;0,0.4)$, $f_Q(q;0.5,0.4)$ and $f_Q(q;1,0.4)$ in Example 11.16.

In the case that $Q_0(x)$ is not stochastic but deterministic, say a function $g(x)$, in [58, Remark 2.3] the authors justified that, if $g(x)$ is smooth and the

equation $g(x) - q = 0$ has n isolated zeros, $x_{j,q}$ $(j = 1, \ldots, n)$, and $g'(x)$ does not vanish at each of those zeros, then

$$f_Q(q; x, t) = \sum_{j=1}^{n} \frac{f_{A(t)}(x - x_{j,q})}{|g'(x_{j,q})|}.$$

Instead of using this equality to obtain the probability distribution of $Q(x, t)$, we can proceed by utilizing (11.13) and the law of large numbers:

$$F_Q(q; x, t) = \mathbb{E}[F_{Q_0}(q; x - A(t))] \approx \frac{1}{M} \sum_{i=1}^{M} F_{Q_0}(q; x - A^{(i)}(t)). \tag{11.16}$$

If $Q_0(x)$ is a deterministic function, $g(x)$, we may write $F_{Q_0}(q; x - A^{(i)}(t))$ above as $H(q - g(x - A^{(i)}(t)))$, where H is the Heaviside function.

The distribution function can also be used for uncertainty quantification for $Q(x, t)$. Indeed, any (absolute and standard) statistical moment of $Q(x, t)$ may be calculated by using the following formulas:

$$\mathbb{E}[|Q(x,t)|^n] = n \int_0^\infty (1 - F_Q(q; x, t)) q^{n-1} \, \mathrm{d}q + n \int_{-\infty}^0 F_Q(q; x, t) |q|^{n-1} \, \mathrm{d}q, \tag{11.17}$$

$$\mathbb{E}[Q(x,t)^n] = n \int_0^\infty (1 - F_Q(q; x, t)) q^{n-1} \, \mathrm{d}q - n \int_{-\infty}^0 F_Q(q; x, t) q^{n-1} \, \mathrm{d}q. \tag{11.18}$$

Example 11.17 Let us solve [58, Example 2.1] with (11.16). Consider

$$Q_0(x) = \begin{cases} 1, & x \leq 0, \\ 0, & x > 0. \end{cases}$$

According to [58, Example 2.1], $Q(x, t)$ has a Bernoulli distribution, with $\mathbb{P}(Q(x, t) = 1) = 1 - F_{A(t)}(x)$. Let us check this result again with $V(t) = V \sim \text{Binomial}(3, 0.7)$. Figure 11.4 plots $F_Q(q; 1, 0.4)$. We observe the distribution function of a Bernoulli distribution, as predicted in [58, Example 2.1]. Moreover, the first jump discontinuity takes on the value $\mathbb{P}(Q(1, 0.4) = 0) = F_{A(0.4)}(1) = 0.657$.

Example 11.18 Consider $Q_0(x) = Bx$ and $V(t) = V$, $B \sim \text{Binomial}(20, 0.5)$ and $V \sim \text{Binomial}(2, 0.7)$ are independent. In Figure 11.5, we plot $F_Q(q; 1, 0.4)$, $F_Q(q; 2, 0.4)$, $F_Q(q; 10, 0.4)$ and $F_Q(q; 20, 0.4)$ via (11.16). We observe that these distribution functions have a shape more complicated than in the previous example. Formulas (11.17) and (11.18) may be used for uncertainty quantification for $Q(x, t)$.

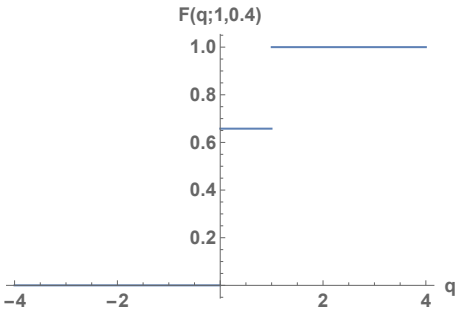

Figure 11.4: Graph of $F_Q(q; 1, 0.4)$ in Example 11.17.

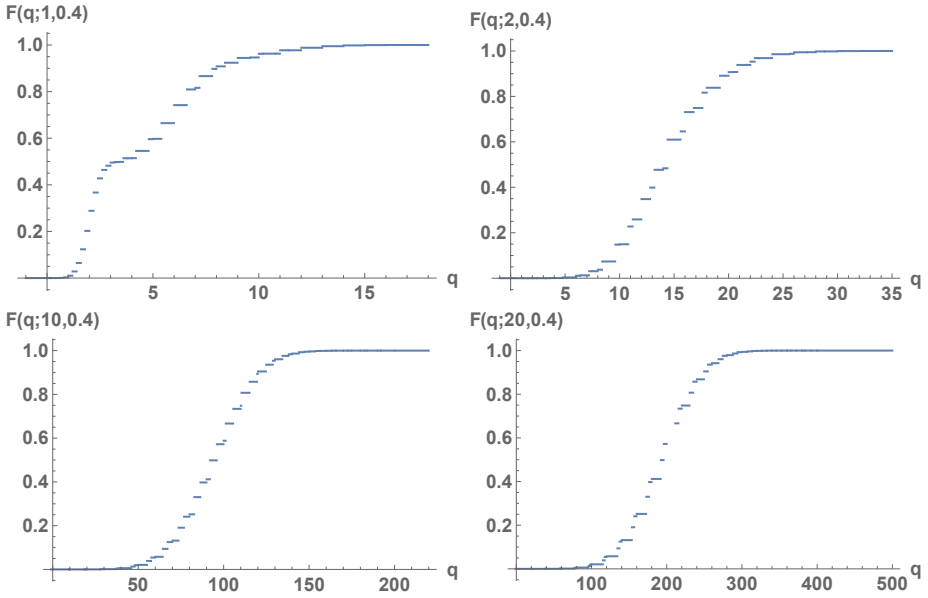

Figure 11.5: Graph of $F_Q(q; 1, 0.4)$, $F_Q(q; 2, 0.4)$, $F_Q(q; 10, 0.4)$ and $F_Q(q; 20, 0.4)$ in Example 11.18.

11.4 Partial differential equation for the density function

In this section, we derive a theoretical partial differential equation for $f_Q(q;x,t)$ in a general setting. In [58, Prop. 3.1], the authors deduced a partial differential equation in the case that $V(t)$ is a Gaussian stochastic process. Thus, this section is a complement of [58, Prop. 3.1].

Theorem 11.19 *Suppose that $Q_0(x)$ is a stochastic process with absolutely continuous law for each $x \in \mathbb{R}$. Let $V(t)$ be any stochastic process, $t \in [0,T]$. Assume that Q_0 and V are independent processes. Then the probability density function of $Q(x,t)$, $f_Q(q;x,t)$, satisfies the following partial differential equation for $x \in \mathbb{R}$ and $t \in [0,T]$:*

$$\frac{\partial}{\partial t} f_Q(q;x,t) + \frac{\partial}{\partial x}\left\{\mathbb{E}[V(t)|Q(x,t)=q] f_Q(q;x,t)\right\} = 0.$$

Proof. By (11.14), $Q(x,t)$ is absolutely continuous with density function given by $f_Q(q;x,t) = \mathbb{E}[f_{Q_0}(q;x-A(t))]$, where $A'(t) = V(t)$. We differentiate with respect to t:

$$\begin{aligned}\frac{\partial}{\partial t} f_Q(q;x,t) =& -\mathbb{E}[V(t)\,\partial_2 f_{Q_0}(q;x-A(t))] \\ =& -\int_{\mathbb{R}^2} v\,\partial_2 f_{Q_0}(q;x-a)\,\mathbb{P}_{(V(t),A(t))}(\mathrm{d}v,\mathrm{d}a). \end{aligned} \tag{11.19}$$

On the other hand, we write

$$\begin{aligned}&\frac{\partial}{\partial x}\left\{\mathbb{E}[V(t)|Q(x,t)=q] f_Q(q;x,t)\right\} \\ &= \frac{\partial}{\partial x}\left\{ f_Q(q;x,t)\int_{\mathbb{R}^2} v\,\mathbb{P}_{(V(t),A(t))|Q(x,t)=q}(\mathrm{d}v,\mathrm{d}a)\right\}. \end{aligned} \tag{11.20}$$

Let us see that

$$\mathbb{P}_{(V(t),A(t))|Q(x,t)=q}(\mathrm{d}v,\mathrm{d}a) = \frac{f_{Q_0}(q;x-a)}{f_Q(q;x,t)}\,\mathbb{P}_{(V(t),A(t))}(\mathrm{d}v,\mathrm{d}a) \tag{11.21}$$

(we can divide by $f_Q(q;x,t)$, since the conditional law $\mathbb{P}_{(V(t),A(t))|Q(x,t)=q}$ is defined for every q except in a set of probability 0 with respect to $\mathbb{P}_{Q(x,t)}$, and $\mathbb{P}_{Q(x,t)}(\{q\in\mathbb{R}:\ f_Q(q;x,t)=0\}) = \int_{\{q:\,f_Q(q;x,t)=0\}} f_Q(q;x,t)\,\mathrm{d}q = 0$). By the Radon-Nikodym Theorem, (11.21) is equivalent to

$$\mathbb{P}(V(t)\in C_1, A(t)\in C_2|Q(x,t)=q) = \int_{C_1\times C_2} \frac{f_{Q_0}(q;x-a)}{f_Q(q;x,t)}\,\mathbb{P}_{(V(t),A(t))}(\mathrm{d}v,\mathrm{d}a), \tag{11.22}$$

for any Borel sets $C_1, C_2 \subseteq \mathbb{R}$. The conditional probability $\mathbb{P}_{(V(t),A(t))|Q(x,t)=q}$ is fully determined by the law of total probability

$$\begin{aligned}&\mathbb{P}(V(t) \in C_1, A(t) \in C_2, Q(x,t) \in B)\\ &= \int_B \mathbb{P}(V(t) \in C_1, A(t) \in C_2 | Q(x,t) = q)\, \mathbb{P}_{Q(x,t)}(\mathrm{d}q)\end{aligned}$$

(B Borel set in $\mathbb{R}$), therefore (11.22) is equivalent to proving that

$$\begin{aligned}&\mathbb{P}(V(t) \in C_1, A(t) \in C_2, Q(x,t) \in B)\\ &= \int_B \int_{C_1 \times C_2} \frac{f_{Q_0}(q; x-a)}{f_Q(q;x,t)}\, \mathbb{P}_{(V(t),A(t))}(\mathrm{d}v, \mathrm{d}a)\, \mathbb{P}_{Q(x,t)}(\mathrm{d}q). \qquad (11.23)\end{aligned}$$

Expression (11.23) is clear, as a consequence of Fubini's Theorem (justified by the non-negativity of the integrand), the independence between Q_0 and (V, A), and the law of total probability:

$$\begin{aligned}&\int_B \int_{C_1 \times C_2} \frac{f_{Q_0}(q; x-a)}{f_Q(q;x,t)}\, \mathbb{P}_{(V(t),A(t))}(\mathrm{d}v, \mathrm{d}a)\, \mathbb{P}_{Q(x,t)}(\mathrm{d}q)\\ &= \int_B \int_{C_1 \times C_2} \frac{f_{Q_0}(q; x-a)}{f_Q(q;x,t)}\, \mathbb{P}_{(V(t),A(t))}(\mathrm{d}v, \mathrm{d}a) f_Q(q;x,t)\, \mathrm{d}q\\ &= \int_{C_1 \times C_2} \int_B f_{Q_0}(q; x-a)\, \mathrm{d}q\, \mathbb{P}_{(V(t),A(t))}(\mathrm{d}v, \mathrm{d}a)\\ &= \int_{C_1 \times C_2} \mathbb{P}(Q_0(x-a) \in B)\, \mathbb{P}_{(V(t),A(t))}(\mathrm{d}v, \mathrm{d}a)\\ &= \int_{C_1 \times C_2} \mathbb{P}(Q(x,t) \in B | V(t) = v, A(t) = a)\, \mathbb{P}_{(V(t),A(t))}(\mathrm{d}v, \mathrm{d}a)\\ &= \mathbb{P}(V(t) \in C_1, A(t) \in C_2, Q(x,t) \in B).\end{aligned}$$

Therefore, (11.21) holds. By substituting (11.21) into (11.20),

$$\begin{aligned}&\frac{\partial}{\partial x} \{\mathbb{E}[V(t) | Q(x,t) = q] f_Q(q;x,t)\}\\ &= \frac{\partial}{\partial x} \left\{ \int_{\mathbb{R}^2} v f_{Q_0}(q; x-a)\, \mathbb{P}_{(V(t),A(t))}(\mathrm{d}v, \mathrm{d}a) \right\}\\ &= \int_{\mathbb{R}^2} v\, \partial_2 f_{Q_0}(q; x-a)\, \mathbb{P}_{(V(t),A(t))}(\mathrm{d}v, \mathrm{d}a). \qquad (11.24)\end{aligned}$$

Finally, by combining (11.19) and (11.24), we derive the partial differential equation for $f_Q(q;x,t)$.

□

11.5 Joint density function of the solution process

In this section, we want to derive the joint density function of the random vector $(Q(x_1,t),Q(x_2,t))$, denoted as $f_Q(q_1,q_2;x_1,x_2,t)$. In [58, Prop. 4.1], the authors derived the formula

$$f_Q(q_1,q_2;x_1,x_2,t)=\int_{\mathbb{R}} f_{Q_0}(q_1,q_2;x_1-a,x_2-a,t)f_{A(t)}(a)\,\mathrm{d}a. \tag{11.25}$$

For the proof, they used the law of total probability by conditioning the distribution function of $(Q(x_1,t),Q(x_2,t))$, $F_Q(q_1,q_2;x_1,x_2,t)$, with respect to the random vector $(X_0,Y_0)=(x_1-A(t),x_2-A(t))$, to obtain $F_Q(q_1,q_2;x_1,x_2,t)=\int_{\mathbb{R}^2}F_{Q_0}(q_1,q_2;x_0,y_0)f_{X_0,Y_0}(x_0,y_0)\,\mathrm{d}x_0\,\mathrm{d}y_0$. By differentiating with respect to q_1 and q_2, the authors expressed the joint density function of $(Q(x_1,t),Q(x_2,t))$ as $f_Q(q_1,q_2;x_1,x_2,t)=\int_{\mathbb{R}^2}f_{Q_0}(q_1,q_2;x_0,y_0)f_{X_0,Y_0}(x_0,y_0)\,\mathrm{d}x_0\,\mathrm{d}y_0$. By utilizing properties of the Heaviside function and the Dirac delta function, the authors expressed the joint density function of (X_0,Y_0), f_{X_0,Y_0}, in terms of the Dirac delta function: $f_{X_0,Y_0}(x_0,y_0)=f_{A(t)}(x_1-x_0)\delta(x_1-x_0-(x_2-y_0))$. Finally, from the properties of the delta function, (11.25) was deduced.

The goal of this section is to derive (11.25) with a much simpler and shorter proof than in [58]. The idea is to use the law of total probability in an appropriate way to straightforwardly obtain (11.25), without making use of the Heaviside and Dirac delta functions. Instead of conditioning with respect to $(X_0,Y_0)=(x_1-A(t),x_2-A(t))$, we condition with respect to $A(t)$. If we denote by $\mathbb{P}_{A(t)}=\mathbb{P}\circ(A(t))^{-1}$ the probability law of $A(t)$ (which may be absolutely continuous, discrete or singular), then

$$\begin{aligned}
F_Q(q_1,q_2;x_1,x_2,t)&=\mathbb{P}(Q(x_1,t)\le q_1,Q(x_2,t)\le q_2)\\
&=\mathbb{P}(Q_0(x_1-A(t))\le q_1,Q_0(x_2-A(t))\le q_2)\\
&=\int_{\mathbb{R}}\mathbb{P}(Q_0(x_1-A(t))\le q_1,Q_0(x_2-A(t))\le q_2|A(t)=a)\,\mathbb{P}_{A(t)}(\mathrm{d}a)\\
&=\int_{\mathbb{R}}\mathbb{P}(Q_0(x_1-a)\le q_1,Q_0(x_2-a)\le q_2|A(t)=a)\,\mathbb{P}_{A(t)}(\mathrm{d}a)\\
&=\int_{\mathbb{R}}\mathbb{P}(Q_0(x_1-a)\le q_1,Q_0(x_2-a)\le q_2)\,\mathbb{P}_{A(t)}(\mathrm{d}a)\\
&=\int_{\mathbb{R}}F_{Q_0}(q_1,q_2;x_1-a,x_2-a)\,\mathbb{P}_{A(t)}(\mathrm{d}a)\\
&=\mathbb{E}[F_{Q_0}(q_1,q_2;x_1-A(t),x_2-A(t))],
\end{aligned}$$

where the independence between the velocity and the initial condition processes, V and Q_0, has been used. As a consequence,

$$\begin{aligned} f_Q(q_1,q_2;x_1,x_2,t) &= \int_{\mathbb{R}} f_{Q_0}(q_1,q_2;x_1-a,x_2-a)\,\mathbb{P}_{A(t)}(\mathrm{d}a) \\ &= \mathbb{E}[f_{Q_0}(q_1,q_2;x_1-A(t),x_2-A(t))]. \end{aligned}$$

If $A(t)$ is an absolutely continuous random variable, then $\mathbb{P}_{A(t)}(\mathrm{d}a) = f_{A(t)}(a)\mathrm{d}a$, so (11.25) has been deduced again.

We finish this section by stating a partial differential equation for the joint probability density function $f_Q(q_1,q_2;x_1,x_2,t)$. We omit the details of the proof, since it is completely analogous to that of Theorem 11.19.

Theorem 11.20 *Suppose that $Q_0(x)$ is a stochastic process such that each random vector $(Q_0(x_1),Q_0(x_2))$ is absolutely continuous for each distinct points $x_1,x_2\in\mathbb{R}$. Let $V(t)$ be any stochastic process, $t\in[0,T]$. Assume that Q_0 and V are independent processes. Then the probability density function of $(Q(x_1,t),Q(x_2,t))$, $f_Q(q_1,q_2;x_1,x_2,t)$, satisfies the following partial differential equation for $x_1\neq x_2\in\mathbb{R}$ and $t\in[0,T]$:*

$$\begin{aligned} &\frac{\partial}{\partial t} f_Q(q_1,q_2;x_1,x_2,t) \\ &+\left(\frac{\partial}{\partial x_1}+\frac{\partial}{\partial x_2}\right)\{\mathbb{E}[V(t)|Q(x_1,t)=q_1,Q(x_2,t)=q_2]f_Q(q_1,q_2;x_1,x_2,t)\}=0. \end{aligned}$$

11.6 Conclusions

In this chapter, we extended the study on the one-dimensional linear advection equation subject to independent stochastic velocity and initial condition performed in [58]. We proved a general version of the random chain rule for the composition of two stochastic processes. This allowed us to find sufficient conditions under which the sample-path solution is a mean square solution as well. By using the law of total probability and the law of large numbers, we showed how to approximate the probability density function or the cumulative distribution function (in the case that the initial condition is not absolutely continuous). Furthermore, we have established the uniqueness of the solution stochastic process, in the mean square sense, to the random one-dimensional linear advection equation under different hypotheses. Our approach does not require the probability distribution of the integral of the stochastic velocity,

but exact or even approximate realizations of it. Thus, non-Gaussian velocity cases can be tackled. This is one of the main results of this contribution, since our approach allows us to treat the non-Gaussian scenario, which is more realistic from a practical standpoint and, to the best of our knowledge, there are not available results in this respect. Several numerical experiments illustrated the theoretical ideas. On the other hand, we deduced a theoretical partial differential equation for the density function of the solution process in a general setting. Our approach could be useful for other random partial differential equations different to the linear advection equation. Finally, a shorter and easier derivation of the joint probability density function of the response process at two spatial points was obtained by applying the law of total probability in a suitable manner.

Acknowledgements

This work has been supported by the Spanish Ministerio de Economía y Competitividad grant MTM2017–89664–P. I acknowledge the doctorate scholarship granted by Programa de Ayudas de Investigación y Desarrollo (PAID), Universitat Politècnica de València.

The main results of this chapter have been published in [16]. This chapter is an international collaboration with the researcher Fábio Antonio Dorini (Department of Mathematics, Federal University of Technology - Paraná 80230-901, Curitiba, PR, Brazil).

Bibliography

[1] E. Allen. *Modeling with Itô Stochastic Differential Equations*. Mathematical Modelling: Theory and Applications. Netherlands: Springer, 2007. ISBN: 978140205953. (Cited on page 1.)

[2] A. Ambrosetti and G. Prodi. *A Primer of Nonlinear Analysis*. Cambridge Studies in Advanced Mathematics. New York: Cambridge University Press, 1993. ISBN: 9780521485739. (Cited on page 104.)

[3] T.M. Apostol. *Mathematical Analysis*. Ed. 2. New York: Pearson, 1976. (Cited on page 25.)

[4] R.B. Ash. *Real Analysis and Probability*. New York: Academic Press, 1972. (Cited on page 102.)

[5] F.M. Asl and A.G. Ulsoy. "Analysis of a system of linear delay differential equations". In: *J. Dyn. Syst.-T. ASME* 125 (2003), pp. 215–223. DOI: 10.1115/1.1568121. (Cited on page 226.)

[6] M. Benhadri and H. Zeghdoudi. "Mean square asymptotic stability in nonlinear stochastic neutral Volterra-Levin equations with Poisson jumps and variable delays". In: *Functiones et Approximatio Commentarii Mathematici* 58.2 (2018), pp. 157–176. DOI: 10.7169/facm/1657. (Cited on page 224.)

[7] A. Berman and R.J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*. New York: SIAM, 1994. (Cited on page 214.)

[8] A. Beskos, A. Jasra, K. Law, R. Tempone, and Y. Zhou. "Multilevel sequential Monte Carlo samplers". In: *Stoch. Proc. Appl.* 127.5 (2017), pp. 1417–1440. DOI: 10.1016/j.spa.2016.08.004. (Cited on page 139.)

[9] P. Billingsley. *Probability and Measure*. Ed. 3. Wiley Series in Probability and Mathematical Statistics. India: Wiley, 1995. (Cited on page 218.)

[10] G.A. Bocharov and F.A. Rihan. "Numerical modelling in biosciences using delay differential equations". In: *J. Comput. Appl. Math.* 125 (2000), pp. 183–199. DOI: 10.1016/S0377-0427(00)00468-4. (Cited on page 224.)

[11] W.E. Boyce. *Probabilistic Methods in Applied Mathematics I*. New York: Academic Press, 1968. (Cited on page 37.)

[12] H. Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Universitext. New York: Springer, 2010. (Cited on pages 104 and 209.)

[13] C.M. Brown. *Differential Equations: A Modeling Approach (Quantitative Applications in the Social Sciences)*; New York: SAGE, Inc., 2007. (Cited on page 1.)

[14] J. Calatayud, B.M. Chen-Charpentier, J.-C. Cortés, and M. Jornet. "Combining polynomial chaos expansions and the random variable transformation technique to approximate the density function of stochastic problems, including some epidemiological models". In: *Symmetry* 11.1 (2019), p. 43. DOI: 10.3390/sym11010043. (Cited on page 103.)

[15] J. Calatayud, J.-C. Cortés, J.A. Díaz, and M. Jornet. "Density function of random differential equations via finite difference schemes: a theoretical analysis of a random diffusion-reaction Poisson-type problem". In: *Stochastics* (2019). DOI: 10.1080/17442508.2019.1645849. (Cited on page 222.)

[16] J. Calatayud, J.-C. Cortés, F.A. Dorini, and M. Jornet. "Extending the study on the linear advection equation subject to stochastic velocity field and initial condition". In: *Math. Comput. Simulat.* (2020). DOI: 10.1016/j.matcom.2019.12.014. (Cited on page 265.)

[17] J. Calatayud, J.-C. Cortés, and M. Jornet. "L^p-calculus approach to the random autonomous linear differential equation with discrete delay". In: *Mediterr. J. Math.* 16.4 (2019), p. 85. DOI: 10.1007/s00009-019-1370-6. (Cited on page 240.)

[18] J. Calatayud, J.-C. Cortés, and M. Jornet. "On the convergence of adaptive gPC for non-linear random difference equations: Theoretical analysis and some practical recommendations". In: *J. Nonlinear Sci. App.*, 11.9 (2018), pp. 1077–1084. DOI: 10.22436/jnsa.011.09.06. (Cited on page 78.)

[19] J. Calatayud, J.-C. Cortés, and M. Jornet. "Uncertainty quantification for random parabolic equations with nonhomogeneous boundary conditions on a bounded domain via the approximation of the probability density function". In: *Math. Method. Appl. Sci.* (2018), pp. 1–19. DOI: 10.1002/mma.5333. (Cited on pages 103 and 222.)

[20] J. Calatayud, J.-C. Cortés, and M. Jornet. "The damped pendulum random differential equation: A comprehensive stochastic analysis via the computation of the probability density function". In: *Physica A* 512 (2018), pp. 261–279. DOI: 10.1016/j.physa.2018.08.024. (Cited on pages 103 and 104.)

[21] J. Calatayud, J.-C. Cortés, and M. Jornet. "Improving the approximation of the first and second order statistics of the response stochastic process to the random Legendre differential equation". In: *Mediterr. J. Math.* 16.3 (2019), p. 68. DOI: 10.1007/s00009-019-1338-6. (Cited on pages 20, 192, and 193.)

[22] J. Calatayud, J.-C. Cortés, and M. Jornet. "Computational uncertainty quantification for random non-autonomous second order linear differential equations via adapted gPC: a comparative case study with random Fröbenius method and Monte Carlo simulation". In: *Open Math.* 16 (2018), pp. 1651–1666. DOI: 10.1515/math-2018-0134. (Cited on page 95.)

[23] J. Calatayud, J.-C. Cortés, and M. Jornet. "On the Legendre differential equation with uncertainties at the regular-singular point 1: $L^p(\Omega)$ random power series solution and approximation of its statistical moments". In: *Computational and Mathematical Methods* 1:e1045 (2019). DOI: 10.1002/cmm4.1045. (Cited on page 205.)

[24] J. Calatayud, J.-C. Cortés, and M. Jornet. "Some notes to extend the study on random non-autonomous second order linear differential equations appearing in mathematical modeling". In: *Mathematical and Computational Applications* 23.4 (2018), pp. 76–89. DOI: 10.3390/mca23040076. (Cited on page 59.)

[25] J. Calatayud, J.-C. Cortés, M. Jornet, and L. Villafuerte. "Random non-autonomous second order linear differential equations: mean square analytic solutions and their statistical properties". In: *Adv. Differ. Equ.* 2018.392 (2018), pp. 1–29. DOI: 10.1186/s13662-018-1848-8. (Cited on page 59.)

[26] J. Calatayud, J.-C. Cortés, M. Jornet, and R.-J. Villanueva. "Computational uncertainty quantification for random time-discrete epidemiological models using adaptive gPC". In: *Math. Method Appl. Sci.* (2018). DOI: 10.1002/mma.5315. (Cited on page 200.)

[27] G. Calbo. *Mean Square Analytic Solutions of Random Linear Models*. Doctoral dissertation. Universitat Politècnica de València, Spain, 2010. DOI: 10.4995/Thesis/10251/8721. (Cited on pages 4 and 63.)

[28] G. Calbo, J.-C. Cortés, and L. Jódar. "Mean square power series solution of random linear differential equations". In: *Comput. Math. Ap-*

plic. 59 (2010), pp. 559–572. DOI: 10.1016/j.camwa.2009.06.007. (Cited on page 63.)

[29] G. Calbo, J.-C. Cortés, and L. Jódar. "Random Hermite differential equations: Mean square power series solutions and statistical properties". In: *Appl. Math. Comput.* 218.7 (2011), pp. 3654–3666. DOI: 0.1016/j.amc.2011.09.008. (Cited on pages 4, 8, 9, 22, 32, 37, 41, 47, 62, 63, 64, 66, and 70.)

[30] G. Calbo, J.-C. Cortés, L. Jódar, and L. Villafuerte. "Analytic stochastic processes solutions of second-order random differential equations". In: *Appl. Math. Lett.* 23 (2010), pp. 1421–1424. DOI: 10.1016/j.aml.2010.07.011. (Cited on page 63.)

[31] G. Calbo, J.-C. Cortés, L. Jódar, and L. Villafuerte. "Solving the random Legendre differential equation: Mean square power series solution and its statistical functions". In: *Comput. Math. Appl.* 61.9 (2011), pp. 2782–2792. DOI: 10.1016/j.camwa.2011.03.045. (Cited on pages 8, 9, 10, 15, 22, 192, and 193.)

[32] T. Caraballo, R. Colucci, and L. Guerrini. "On a predator prey model with nonlinear harvesting and distributed delay". In: *Commun. Pur. Appl. Anal.* 17.6 (2018), pp. 2703–2727. DOI: 10.3934/cpaa.2018128. (Cited on page 224.)

[33] C.M. Carracedo and M.A.S. Alix. *Introducción a las Ecuaciones Diferenciales Ordinarias*. Barcelona: Reverté, 1991. ISBN: 9788429150438. (written in Spanish). (Cited on pages 4 and 193.)

[34] B.-M. Chen-Charpentier, J.-C. Cortés, J.-A. Licea, J.-V. Romero, M.-D. Roselló, F.-J. Santonja, and R.-J. Villanueva. "Constructing adaptive generalized polynomial chaos method to measure the uncertainty in continuous models: A computational approach". In: *Math. Comput. Simulat.* 109 (2015), pp. 113–129. DOI: 10.1016/j.matcom.2014.09.002. (Cited on pages 3, 76, 77, 78, and 80.)

[35] B.-M. Chen-Charpentier, J.-C. Cortés, J.-V. Romero, and M.-D. Roselló. "Some recommendations for applying gPC (generalized polynomial chaos) to modeling: An analysis through the Airy random differential equation". In: *Appl. Math. Comput.* 219.9 (2013), pp. 4208–4218. DOI: 10.1016/j.amc.2012.11.007. (Cited on page 76.)

[36] B.-M. Chen-Charpentier, J.-C. Cortés, J.-V. Romero, and M.-D. Roselló. "Do the generalized polynomial chaos and Fröbenius methods retain the statistical moments of random differential equations?". In: *Appl. Math. Lett.* 26.5 (2013), pp. 553–558. DOI: 10.1016/j.aml.2012.12.013. (Cited on page 76.)

[37] B.M. Chen-Charpentier and I. Diakite. "A mathematical model of bone remodeling with delays". In: *J. Comput. Appl. Math.* 291 (2016), pp. 76–84. DOI: 10.1016/j.cam.2017.01.005. DOI: 10.1016/j.cam.2014.11.025. (Cited on page 224.)

[38] M.C. Casabán, J.-C. Cortés, J.V. Romero, and M.D. Roselló. "Solving random homogeneous linear second-order differential equations: A full probabilistic description". In: *Mediterr. J. Math.* 13.6 (2016), pp. 3817–3836. DOI: 10.1007/s00009-016-0716-6. (Cited on pages 22 and 98.)

[39] M.C. Casabán, J.-C. Cortés, A. Navarro-Quiles, J.V. Romero, M.D. Roselló, and R.-J. Villanueva. "Computing probabilistic solutions of the Bernoulli random differential equation". In: *J. Comput. Appl. Math.* 309 (2017), pp. 396–407. DOI: 10.1016/j.cam.2016.02.034. (Cited on pages 53 and 54.)

[40] B.M. Chen-Charpentier and D. Stanescu. "Epidemic models with random coefficients". In: *Math. Comput. Model.* 52.7–8 (2010), pp. 1004–1010. DOI: 10.1016/j.mcm.2010.01.014. (Cited on page 76.)

[41] C. Chicone. *Ordinary Differential Equations with Applications*. Vol 34. Texts in Applied Mathematics. New York: Springer-Verlag, 2006. ISBN: 9780387307695. (Cited on pages 142 and 197.)

[42] A. Cliffe, M.B. Giles, R. Scheichl, and A. Teckentrup. "Multilevel Monte Carlo Methods and applications to elliptic PDEs with random coefficients". In: *Computing and Visualization in Science* 14.1 (2011), pp. 3–15. DOI: 10.1007/s00791-011-0160-x. (Cited on page 139.)

[43] J.-C. Cortés, L. Jódar, J. Camacho, and L. Villafuerte. "Random Airy type differential equations: Mean square exact and numerical solutions". In: *Comput. Math. Appl.* 60.5 (2010), pp. 1237–1244. DOI: 10.1016/j.camwa.2010.05.046. (Cited on pages 4, 8, 9, 22, 32, 37, 41, 42, 62, 63, 66, 86, 87, and 89.)

[44] J.-C. Cortés, L. Jódar, R. Company, and L. Villafuerte. "Laguerre random polynomials: definition, differential and statistical properties". In: *Utilitas Mathematica* 98 (2015), pp. 283–295. (Cited on pages 22, 62, 63, 66, and 68.)

[45] J.-C. Cortés, L. Jódar, M.D. Roselló, and L. Villafuerte. "Solving initial and two-point boundary value linear random differential equations: A mean square approach". In: *Appl. Math. Comput.* 219.4 (2012), pp. 2204–2211. DOI: 10.1016/j.amc.2012.08.066. (Cited on pages 246, 247, 252, and 253.)

[46] J.-C. Cortés, L. Jódar, and L. Villafuerte. "Mean square solution of Bessel differential equation with uncertainties". In: *J. Comput. Appl. Math.* 309.1

(2017), pp. 383–395. DOI: 10.1016/j.cam.2016.01.034- (Cited on pages 8 and 22.)

[47] J.-C. Cortés, A. Navarro-Quiles, J.V. Romero, and M.D. Roselló. "Probabilistic solution of random autonomous first-order linear systems of ordinary differential equations". In: *Rom. Rep. Phys.* 68 (2016), pp. 1397–1406. (Cited on pages 77 and 217.)

[48] J.-C. Cortés, A. Navarro-Quiles, J.V. Romero, and M.D. Roselló. "Solving second-order linear differential equations with random analytic coefficients about ordinary points: A full probabilistic solution by the first probability density function". In: *Appl. Math. Comput.* 331 (2018), pp. 33–45. DOI: 10.1016/j.amc.2018.02.051. (Cited on pages 98, 102, and 103.)

[49] J.-C. Cortés, J.V. Romero, M.D. Roselló, F.J. Santonja, and R.J. Villanueva. "Solving continuous models with dependent uncertainty: a computational approach". In: *Abstr. Appl. Anal.* 2013 (2013). DOI: 10.1155/2013/983839. (Cited on pages 4, 76, 77, 78, 80, 82, and 200.)

[50] J.-C. Cortés, J.-V. Romero, M.-D. Roselló, and R.-J. Villanueva. "Dealing with dependent uncertainty in modelling: a comparative study case through the Airy equation". In: *Abstr. Appl. Anal.* 2013 (2013). DOI: 10.1155/2013/279642. (Cited on page 76.)

[51] J.-C. Cortés, J.-V. Romero, M.-D. Roselló, and R.-J. Villanueva. "Improving adaptive generalized polynomial chaos method to solve nonlinear random differential equations by the random variable transformation technique". In: *Commun. Nonlinear Sci. Numer. Simul.* 50 (2017), pp. 1–15. DOI: 10.1016/j.cnsns.2017.02.011. (Cited on pages 3, 76, 77, and 78.)

[52] J.-C. Cortés, P. Sevilla-Peris, and L. Jódar. "Analytic-numerical approximating processes of diffusion equation with data uncertainty". In: *Comput. Math. Appl.* 49.7-8 (2005), pp. 1255–1266. DOI: 10.1016/j.camwa.2004.05.015. (Cited on page 23.)

[53] J.-C. Cortés, L. Villafuerte, and C. Burgos. "A mean square chain rule and its application in solving the random Chebyshev differential equation". In: *Mediterr. J. Math.* 14.1 (2017), p. 35. (Cited on pages 244 and 245.)

[54] G. Dagan. *Flow and Transport in Porous Formations*. Heidelberg: Springer, 1989. (Cited on page 242.)

[55] A.C. Davison and D.V. Hinkley. *Bootstrap Methods and Their Application*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1997. ISBN 0-521-57391-2. (Cited on page 159.)

[56] J.L. Doob. *Stochastic Processes*. John Wiley & Sons, 1953. (Cited on page 33.)

[57] F.A. Dorini, M.S. Cecconello, and M.B. Dorini. "On the logistic equation subject to uncertainties in the environmental carrying capacity and initial population density". In: *Commun. Nonlinear Sci. Numer. Simul.* 33 (2016), pp. 160–173. DOI: 10.1016/j.cnsns.2015.09.009. (Cited on pages 22 and 208.)

[58] F.A. Dorini and M.C.C. Cunha. "On the linear advection equation subject to random velocity fields". In: *Math. Comput. Simulat.* 82.4 (2011), pp. 679–690. DOI: 10.1016/j.matcom.2011.10.008. (Cited on pages 243, 247, 254, 255, 258, 259, 261, 263, and 264.)

[59] F. Dorini and M. Cunha. "Statistical moments of the random linear transport equation". In: *J. Comput. Phys.* 227.19 (2008), pp. 8541–8550. DOI: 10.1016/j.jcp.2008.06.002. (Cited on page 22.)

[60] Y. Driver. *Ordinary and Delay Differential Equations*. Applied Mathematical Science Series. New York: Springer, 1977. (Cited on page 223.)

[61] B. Efron and R. Tibshirani. *An Introduction to the Bootstrap*. CRC Press, 1994. (Cited on page 159.)

[62] M.A. El-Tawil. "The approximate solutions of some stochastic differential equations using transformations". In: *Appl. Math. Comput.* 164 (2005), pp. 167–178. DOI: 10.1016/j.amc.2004.04.062. (Cited on pages 103 and 208.)

[63] M. El-Tawil, W. El-Tahhan, and A. Hussein. "Using FEM-RVT technique for solving a randomly excited ordinary differential equation with a random operator". In: *Appl. Math. Comput.* 187.2 (2007), pp. 856–867. DOI: 10.1016/j.amc.2006.08.164. (Cited on page 103.)

[64] S.A. El-Wakil, M. Sallah, and A.M. El-Hanbaly. "Random variable transformation for generalized stochastic radiative transfer in finite participating slab media". In: *Physica A* 435 (2015), pp. 66–79. DOI: 10.1016/j.physa.2015.04.033. (Cited on page 243.)

[65] T. Erneux. *Applied Delay Differential Equations*. Surveys and Tutorials in the Applied Mathematical Sciences Series. New York: Springer, 2009. (Cited on page 224.)

[66] O.G. Ernst, A. Mugler, H.J. Starkloff, and E. Ullmann. "On the convergence of generalized polynomial chaos expansions". In: *ESAIM-Math. Model. Num.* 46.2 (2012), pp. 317–339. DOI: 10.1051/m2an/2011045. (Cited on pages 68 and 78.)

[67] L.C. Evans. *Partial Differential Equations*. Orient Blackswan, 2014. (Cited on page 215.)

[68] L.C. Evans. *An Introduction to Stochastic Differential Equations*. American Mathematical Society, 2012. ISBN: 9781470410544. (Cited on page 1.)

[69] W. Fang and M.B. Giles. "Multilevel Monte Carlo method for ergodic SDEs without contractivity". In: *J. Math. Anal. Appl.* 476.1 (2019), pp. 149–176. DOI: 10.1016/j.jmaa.2018.12.032. (Cited on page 139.)

[70] G. Fishman. *Monte Carlo: Concepts, Algorithms, and Applications*. Springer Series in Operations Research and Financial Engineering. New York: Springer Science & Business Media, 2013. (Cited on pages 3, 15, and 199.)

[71] E. Fridman and L. Shaikhet. "Delay-induced stability of vector second-order systems via simple Lyapunov functionals". In: *Automatica* 74 (2016), pp. 288–296. DOI: 10.1016/j.automatica.2016. 07.034. (Cited on page 224.)

[72] M.J. Garrido-Atienza, A. Ogrowsky, and B. Schmalfuss. "Random differential equations with random delays". In: *Stoch. Dynam.* 11.2–3 (2011), pp. 369–388. DOI: 10.1142/S0219493711003358. (Cited on page 225.)

[73] K.O. Geddes, S.R. Czapor, and G. Labahn. *Algorithms for Computer Algebra*. New York: Springer, 1992. ISBN 9780792392590. (Cited on page 109.)

[74] M. Gerritsma, J.B. Van der Steen, P. Vos, and G. Karniadakis. "Time-dependent generalized polynomial chaos". In: *J. Comput. Phys.* 229.22 (2010), pp. 8333–8363. DOI: 10.1016/j.jcp.2010.07.020. (Cited on page 100.)

[75] M.B. Giles. "Multilevel Monte Carlo path simulation". In: *Oper. Res.* 56.3 (2008), pp. 607–617. DOI: 10.1287/opre.1070.0496. (Cited on pages 134, 139, and 166.)

[76] M.B. Giles. "Multilevel Monte Carlo methods". In: *Acta Numer.* 24 (2015), pp. 259–328. DOI: 10.1017/S09624929. (Cited on pages 134, 139, and 166.)

[77] A.K. Golmankhaneh, N.A. Porghoveh, and D. Baleanu. "Mean square solutions of second-order random differential equations by using homotopy analysis method". In: *Rom. Rep. Phys.* 65.2 (2013), pp. 350–362. (Cited on pages 22 and 192.)

[78] B. Goodwine. *Engineering Differential Equations. Theory and Applications*. New York: Springer, 2011. ISBN: 9781441979186. (Cited on page 1.)

[79] D. Gottlieb and D. Xiu. "Galerkin method for wave equations with uncertain coefficients". In: *Commun. Comput. Phys.* 3.2 (2008), pp. 505–518. (Cited on page 78.)

[80] G. Grimmett and D. Stirzaker. *Probability and Random Processes*. Clarendon Press, Oxford, 2000. (Cited on page 36.)

[81] J.K. Hale. *Ordinary Differential Equations*. Ed. 2. Malabar: Robert E. Krieger Publishing Company, 1980. (Cited on page 79.)

[82] L. Harding and M. Neamtu. "A dynamic model of unemployment with migration and delayed policy intervention". In: *Comput. Econ.* 51.3 (2018), pp. 427–462. DOI: 10.1007/s10614-016-9610-3. (Cited on page 224.)

[83] F. Hartung and M. Pituk. *Recent Advances in Delay Differential and Differences Equations*. Berlin Heidelberg: Springer-Verlag, 2014. (Cited on page 224.)

[84] E. Hellinger. "Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen". In: *Journal für die reine und angewandte Mathematik* 136 (1909), pp. 210–271. DOI: 10.1515/crll.1909.136.210. (Cited on page 104.)

[85] D. Henderson and P. Plaschko. *Stochastic Differential Equations in Science and Engineering*, Singapore: World Scientific, 2006. (Cited on pages 37, 85, and 86.)

[86] H.W. Hethcote. "The mathematics of infectious diseases". In: *SIAM Rev.* 42.4 (2000), pp. 599–653. DOI: 10.1137/S0036144500371907. (Cited on page 1.)

[87] A. Hussein and M.M. Selim. "Solution of the stochastic radiative transfer equation with Rayleigh scattering using RVT technique". In: *Appl. Math. Comput.* 218 (2012), pp. 7193–7203. DOI: 10.1016/j.amc.2011.12.088. (Cited on pages 22 and 208.)

[88] A. Hussein and M.M. Selim. "Solution of the stochastic generalized shallow-water wave equation using RVT technique". In: *Eur. Phys. J. Plus* 130.249 (2015). DOI: 10.1140/epjp/i2015-15249-3. (Cited on page 243.)

[89] A. Hussein and M.M. Selim. "Solution of the stochastic transport equation of neutral particles with anisotropic scattering using RVT technique". In: *Appl. Math. Comput.* 213.1 (2009), pp. 250–261. DOI: 10.1016/j.amc.2009.03.016. (Cited on pages 22 and 243.)

[90] A. Hussein and M.M. Selim. "A developed solution of the stochastic Milne problem using probabilistic transformations". In: *Appl. Math. Comput.* 216.10 (2009), pp. 2910–2919. DOI: 10.1016/j.amc.2010.04.003. (Cited on page 22.)

[91] M. Jackson and B.M. Chen-Charpentier. "Modeling plant virus propagation with delays". In: *J. Comput. Appl. Math.* 309 (2017), pp. 611–621. DOI: 0.1016/j.cam.2016.04.024. DOI: 10.1016/j.cam.2016.04.024. (Cited on page 224.)

[92] M. Jagadeesan. “Simple analysis of sparse, sign-consistent JL”. (2017) arXiv:1708.02966. (Cited on page 67.)

[93] M. Jardak, C.H. Su, and G.E. Karniadakis. “Spectral polynomial chaos solutions of the stochastic advection equation”. In: *J. Sci. Comput.* 17 (2002), pp. 319–338. (Cited on page 242.)

[94] M. Jornet, J. Calatayud, O.P. Le Maître, and J.-C. Cortés. “Second order linear differential equations with analytic uncertainties: stochastic analysis via the computation of the probability density function”. In: *J. Comput. Appl. Math.* (2020). (Cited on page 135.)

[95] J. Jost. *Partial Differential Equations*. Ed. 2. Graduate Texts in Mathematics. New York: Springer, 2007. (Cited on page 209.)

[96] G. Kallianpur. *Stochastic Filtering Theory*. New York: Springer, 1980. (Cited on page 219.)

[97] M. Khodabin, K. Maleknejad, M. Rostami, and M. Nouri. “Numerical solution of stochastic differential equations by second order Runge-Kutta methods”. In: *Math. Comput. Model.* 53.9–10 (2011), pp. 1910–1920. DOI: 10.1016/j.mcm.2011.01.018. (Cited on page 22.)

[98] M. Khodabin and M. Rostami. “Mean square numerical solution of stochastic differential equations by fourth order Runge-Kutta method and its application in the electric circuits with noise”. In: *Adv. Differ. Equ.* 2015.1 (2015), p. 62. (Cited on page 22.)

[99] A.R. Khudair, A.A. Ameen, and S.L. Khalaf. “Mean square solutions of second-order random differential equations by using variational iteration method”. In: *Appl. Math. Sci.* 5 (2011), pp. 2505–2519. (Cited on page 22.)

[100] A.R. Khudair, A.A. Ameen, and S.L. Khalaf. “Mean square solutions of second-order random differential equations by using Adomian decomposition method”. In: *Appl. Math. Sci.* 51.5 (2011), pp. 2521–2535. (Cited on page 22.)

[101] A.R. Khudair, S.A.M. Haddad, and S.L. Khalaf. “Mean square solutions of second-order random differential equations by using the differential transformation method”. In: *Open Journal of Applied Sciences* 6 (2016), pp. 287–297. (Cited on page 22.)

[102] D.Y. Khusainov, A.F. Ivanov, and I.V. Kovarzh. “Solution of one heat equation with delay”. In: *Nonlinear Oscil.* 12 (2009), pp. 260–282. DOI: 10.1007/s11072-009-0075-3. (Cited on page 226.)

[103] D.Y. Khusainov and M. Pokojovy. “Solving the linear 1d thermoelasticity equations with pure delay”. In: *International Journal of Mathematics and Mathematical Sciences* 2015 (2015), pp. 1–11. DOI: 10.1155/2015/479267. (Cited on page 237.)

[104] A. Klenke. *Probability Theory: A Comprehensive Course*. Ed. 2. Universitext. London: Springer Science & Business Media, 2013. (Cited on page 106.)

[105] P.L. Krapivsky, J.L. Luck, and K. Mallick. "On stochastic differential equations with random delay". In: *Journal of Statistical Mechanics: Theory and Experiment* (2011). DOI: 10.1088/1742-5468/2011/10/P10008. (Cited on page 225.)

[106] E. Kreyszig. *Advanced Engineering Mathematics*. Ed. 4. New York: John Wiley & Sons, 1979. (Cited on page 192.)

[107] Y. Kuang. *Delay Differential Equations: with Applications in Population Dynamics*. Academic Press, 2012. (Cited on page 224.)

[108] Y.N. Kyrychko and S.J. Hogan. "On the use of delay equations in engineering applications". In: *J. Vib. Control* 16.7–8 (2017), pp. 943–960. DOI: 10.1177/1077546309341100. (Cited on pages 224 and 226.)

[109] D. Lamberton and B. Lapeyre. *Introduction to Stochastic Calculus Applied to Finance*. Chapman and Hall/CRC, 2011. (Cited on page 1.)

[110] L. Lang. *Undergraduate Analysis*. Ed. 2. Undergraduate Texts in Mathematics. New York: Springer, 1997. (Cited on page 192.)

[111] O.P Le Maître, L. Mathelin, O.M. Knio, and Y.M. Hussaini. "Asynchronous time integration for polynomial Chaos expansions of uncertain periodic dynamics". In: *Discret. Contin. Dyn. S.* 28.1 (2010), pp. 199–226. DOI: 10.3934/dcds.2010.28.199. (Cited on page 134.)

[112] E. Lesaffre and A.B. Lawson. *Bayesian Biostatistics*. Statistics in Practice. New York: Wiley, 2012. (Cited on page 54.)

[113] R.J. LeVeque. *Numerical Methods for Conservation Laws*. Berlin: Birkhäuser, 1992. (Cited on page 243.)

[114] S.A. Levin, T.G. Hallam, and L.J. Gross. *Applied Mathematical Ecology*. Biomathematics. Berlin Heidelberg: Springer Science & Business Media, 2012. (Cited on page 1.)

[115] J.A. Licea, L. Villafuerte, and B.M. Chen-Charpentier. "Analytic and numerical solutions of a Riccati differential equation with random coefficients". In: *J. Comput. Appl. Math.* 239 (2013), pp. 208–219. DOI: 10.1016/j.cam.2012.09.040. (Cited on pages 22 and 192.)

[116] G.D. Lin. "Recent developments on the moment problem". In: *Journal of Statistical Distributions and Applications* 4.1 (2017). DOI: 10.1186/s40488-017-0059-2. (Cited on page 68.)

[117] L. Liu and T. Caraballo. "Analysis of a stochastic 2D-Navier-Stokes model with infinite delay". In: *J. Dyn. Differ. Equ.* (2018), pp. 1–26. DOI: 10.1007/s10884-018-9703-x. (Cited on page 224.)

[118] S. Liu, A. Debbouche, and J.R. Wang. "Fuzzy delay differential equations. On the iterative learning control for stochastic impulsive differential equations with randomly varying trial lengths". In: *J. Comput. Appl. Math.* 312 (2017), pp. 47–57. DOI: 10.1016/j.cam.2015.10.028. (Cited on page 225.)

[119] G.J. Lord, C.E. Powell, and T. Shardlow. *An Introduction to Computational Stochastic PDEs*. New York: Cambridge University Press, 2014. (Cited on pages 79, 93, 220, 255, and 258.)

[120] V. Lupulescu and U. Abbas. "Fuzzy delay differential equations". In: *Fuzzy Optim. Decis. Ma.* 11.1 (2012), pp. 91–111. DOI: 10.1007/s10700-011-9112-7. (Cited on page 225.)

[121] V. Lupulescu, D. O'Regan, and G.U. Rahman. "Existence results for random fractional differential equations". In: *Opuscula Mathematica* 34.4 (2014), pp. 813–825. DOI: 10.7494/OpMath.2014.34.4.813. (Cited on page 192.)

[122] A. Matsumoto and F. Szidarovszky. *Delay Differential Nonlinear Economic Models*. In Nonlinear Dynamics in Economics, Finance and the Social Sciences, pp. 195–214. Berlin Heidelberg: Springer-Verlag, 2010. (Cited on page 224.)

[123] M.E. Mincsovics. "Stability of one-step and linear multistep methods - a matrix technique approach". In: *Electron. J. Qual. Theo.* 2016 (2016), pp. 1–10. (Cited on pages 210 and 214.)

[124] T. Neckel and F. Rupp. *Random Differential Equations in Scientific Computing*. München, Germany: Walter de Gruyter, 2013. (Cited on page 2.)

[125] L. Norman, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions*. Vol. 1. Wiley, 1994. (Cited on pages 12 and 198.)

[126] K. Nouri and H. Ranjbar. "Improved Euler-Maruyama method for numerical solution of the Itô stochastic differential systems by composite previous-current-step idea". In: *Mediterr. J. Math.* 15.140 (2018). DOI: 10.1007/s00009-018-1187-8. (Cited on page 224.)

[127] K. Nouri and H. Ranjbar. "Mean square convergence of the numerical solution of random differential equations". In: *Mediterr. J. Math.* 12.3 (2015), pp. 1123–1140. DOI: 10.1007/s00009-014-0452-8. (Cited on pages 22, 192, and 225.)

[128] D. Nualart and M. Sanz. *Curs de Probabilitats*. PPU, Barcelona, 1990. (written in Catalan). (Cited on page 102.)

[129] B. Øksendal. *Stochastic Differential Equations. An Introduction with Applications*. Stochastic Modelling and Applied Probability 23. Heidelberg and New York: Springer-Verlag, 2003. (Cited on pages 1 and 224.)

[130] A. Papoulis and U. Pillai. *Probability, Random Variables and Stochastic Processes*. Ed. 4. New York: McGraw-Hill Europe, 2002. (Cited on page 243.)

[131] M. Pisaroni, F.B. Nobile, and P. Leyland. "A Continuation Multi Level Monte Carlo method for uncertainty quantification in compressible inviscid aerodynamics". In: *Comput. Method. Appl. M.* 326.C (2017), pp. 20–50. DOI: 10.1016/j.cma.2017.07.030. (Cited on page 139.)

[132] L.I. Piterbarg and A.G. Ostrovskii. *Advection and Diffusion in Random Media*. Dordrecht: Kluwer Academic Publishers, 1997. (Cited on page 242.)

[133] M.C. Reed. *Fundamental Ideas of Analysis*. New York: John Wiley & Sons, 1998. (Cited on page 192.)

[134] A.C. Rencher and G.B. Schaalje. *Linear Models in Statistics*. Ed. 2. New York: Wiley, 2008. (Cited on page 57.)

[135] W. Rudin. *Principles of Mathematical Analysis*. Ed. 3. Singapore: McGraw-Hill Education, 1976. (Cited on pages 213 and 216.)

[136] T.L. Saaty. *Modern Nonlinear Equations*. New York: Dover Publications Inc., 2003. ISBN: 9780486642321. (Cited on page 2.)

[137] S. Salsa. *Partial Differential Equations in Action, From Modelling to Theory*. Universitext. Milan: Springer, 2010. (Cited on pages 208 and 209.)

[138] A. Saltelli. "Making best use of model evaluations to compute sensitivity indices". In: *Comput. Phys. Commun.* 145 (2002), pp. 280–97. (Cited on page 156.)

[139] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola. "Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index". In: *Comput. Phys. Commun.* 181.2 (2010), pp. 259–270. DOI: 10.1016/j.cpc.2009.09.018. (Cited on page 156.)

[140] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. *Global Sensitivity Analysis. The Primer*. John Wiley & Sons, 2008. (Cited on page 152.)

[141] F. Santonja and L. Shaikhet. "Probabilistic stability analysis of social obesity epidemic by a delayed stochastic model". In: *Nonlinear Anal.: Real* 17 (2014), pp. 114–125. DOI: 10.1016/j.nonrwa.2013.10.010. (Cited on page 224.)

[142] F. Santonja and L. Shaikhet. "Analysing social epidemics by delayed stochastic models". In: *Discrete Dyn. Nat. Soc.* ID 530472 (2012), 13 pages. DOI: 10.1155/2012/530472. (Cited on page 224.)

[143] L.T. Santos, F.A. Dorini, and M.C.C. Cunha. "The probability density function to the random linear transport equation". In: *Appl. Math. Comput.* 216.5 (2010), pp. 1524–1530. DOI: 10.1016/j.amc.2010.03.001. (Cited on page 22.)

[144] H. Scheffé. "A Useful Convergence Theorem for Probability Distributions". In: *Ann. Math. Stat.* 18.3 (1947), pp. 434–438. DOI: 10.1214/aoms/1177730390. (Cited on page 103.)

[145] W.E. Schiesser. *Differential Equation Analysis in Biomedical Science and Engineering: Ordinary Differential Equation Applications with R*. New York: John Wiley & Sons Inc., 2014. (Cited on page 1.)

[146] G.A.F. Seber and C.J. Wild. *Nonlinear Regression*. Cambridge Texts in Applied Mathematics. New York: Wiley, 2003. (Cited on page 53.)

[147] L. Shaikhet. *Lyapunov Functionals and Stability of Stochastic Functional Differential Equations*. New York: Springer, 2013. (Cited on page 224.)

[148] L. Shaikhet. "Stability of equilibrium states of a nonlinear delay differential equation with stochastic perturbations". In: *Int. J. Robust Nonlin.* 27.6 (2016), pp. 915–924. DOI: 10.1002/rnc.3605. (Cited on page 224.)

[149] L. Shaikhet. "About some asymptotic properties of solution of stochastic delay differential equation with a logarithmic nonlinearity". In: *Functional Differential Equations* 4.1–2 (2017), pp. 57–67. (Cited on page 224.)

[150] W. Shi and C. Zhang. "Generalized polynomial chaos for nonlinear random delay differential equations". In: *Appl. Numer. Math.* 115 (2017), pp. 16–31. DOI: 10.1016/j.apnum.2016.12.004. (Cited on pages 78 and 225.)

[151] W. Shi and C. Zhang. "Error analysis of generalized polynomial chaos for nonlinear random ordinary differential equations". *Appl. Numer. Math.*, 62.12 (2012), pp. 1954–1964. DOI: 10.1016/j.apnum.2012.08.007. (Cited on page 78.)

[152] M. Shvidler and K. Karasaki. "Exact averaging of stochastic equations for transport in random velocity field". In: *Transp. Porous Media* 50 (2003), pp. 223–241. (Cited on page 242.)

[153] M. Shvidler and K. Karasaki. "Probability density functions for solute transport in random fields". In: *Transp. Porous Media*, 50 (2003), pp. 243–266. (Cited on page 242.)

[154] H. Slama, N.A. El-Bedwhey, A. El-Depsy, and M.M. Selim. "Solution of the finite Milne problem in stochastic media with RVT Technique". In: *Eur. Phys. J. Plus* 132.505 (2017). (Cited on page 243.)

[155] G.D. Smith. *Numerical Solution of Partial Differential Equations: Finite Difference Methods*. Ed. 3. New York: Oxford University Press, 1986. (Cited on pages 212 and 214.)

[156] H. Smith. *An Introduction to Delay Differential Equations with Applications to the Life Sciences*. Texts in Applied Mathematics. New York: Springer, 2011. (Cited on page 223.)

[157] R.C. Smith. *Uncertainty Quantification. Theory, Implementation and Applications*. Philadelphia: SIAM, 2014. (Cited on pages 2, 54, 57, and 224.)

[158] I.M. Sobol. "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates". In: *Math. Comput. Simulat.* 55.1–3 (2001), pp. 271–280. DOI: 10.1016/S0378-4754(00)00270-6. (Cited on page 152.)

[159] I.M. Sobol', S. Tarantola, D. Gatelli, S. Kucherenko, and W. Mauntz. "Estimating the approximation error when fixing unessential factors in global sensitivity analysis". In: *Reliab. Eng. Syst. Safe.* 92.7 (2007), pp. 957–960. DOI: 10.1016/j.ress.2006.07.001. (Cited on page 156.)

[160] T.T. Soong. *Random Differential Equations in Science and Engineering*. New York: Academic Press, 1973. (Cited on pages 2, 8, 12, 13, 30, 31, 35, 37, 62, 79, 86, 197, 198, 199, 208, 224, 225, 228, 229, 234, 243, 247, 248, 249, 250, 253, and 254.)

[161] J.L. Strand. "Random Ordinary Differential Equations". In: *J. Differ. Equations* 7 (1970), pp. 538–553. DOI: 10.1016/0022-0396(70)90100-2. (Cited on pages 2, 12, 32, 33, 34, 79, 230, and 236.)

[162] L. Tenorio. *An Introduction to Data Analysis and Uncertainty Quantification for Inverse Problems*. Vol. 3. SIAM, 2017. (Cited on page 103.)

[163] M.M. Vainberg. *Variational Methods for the Study of Nonlinear Operators*. Holden-Day, 1964. (Cited on page 104.)

[164] O. Vallée and M. Soares. *Airy Functions and Applications to Physics*. London: Imperial College Press, 2004. (Cited on page 41.)

[165] A.W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. New York: Cambridge University Press, 2000. (Cited on pages 104, 108, and 217.)

[166] D. Venturi, D.M. Tartakovsky, A. M. Tartakovsky, and G. E. Karniadakis. "Exact PDF equations and closure approximations for advective-reactive transport". In: *J. Comput. Phys.* 243 (2013), pp. 323–343. DOI: 10.1016/j.jcp.2013.03.001. (Cited on page 242.)

[167] L. Villafuerte, C.A. Braumann, J.-C. Cortés, and L. Jódar. "Random differential operational calculus: theory and applications". In: *Comput. Math. Appl.* 59.1 (2010), pp. 115–125. DOI: 10.1016/j.camwa.2009.08.061. (Cited on pages 2, 8, 211, 227, 228, 243, 244, 245, 246, 247, 248, 250, 252, and 253.)

[168] L. Villafuerte and B.M. Chen-Charpentier. "A random differential transform method: Theory and applications". In: *Appl. Math. Lett.* 25.10 (2012), pp. 1490–1494. DOI: 10.1016/j.aml.2011.12.033. (Cited on pages 22, 34, 59, and 192.)

[169] N. Wiener. "The homogeneous chaos". In: *Am. J. Math.* 60.4 (1938), pp. 897–936. DOI: 10.2307/2371268. (Cited on page 3.)

[170] D. Williams. *Probability with Martingales*. New York: Cambridge University Press, 1991. (Cited on pages 103 and 216.)

[171] M.M.R. Williams. "Polynomial chaos functions and stochastic differential equations". In: *Ann. Nucl. Energy* 33.9 (2006), pp. 774–785. DOI: 10.1016/j.anucene.2006.04.005. (Cited on page 76.)

[172] D. Xiu. *Numerical Methods for Stochastic Computations. A Spectral Method Approach*. Cambridge Texts in Applied Mathematics. New York: Princeton University Press, 2010. (Cited on pages 3, 38, 68, 76, 77, 78, 79, 85, 103, 109, and 154.)

[173] D. Xiu and G. Karniadakis. "The Wiener-Askey polynomial chaos for stochastic differential equations". In: *J. Sci. Comput.* 24 (2002), pp. 619–644. DOI: 10.1137/S1064827501387826. (Cited on pages 3, 15, 76, and 78.)

[174] D. Zhang. *Stochastic Methods for Flow in Porous Media - Coping with Uncertainties*. San Diego: Academic Press, 2002. (Cited on page 242.)

[175] T. Zhou. "A stochastic collocation method for delay differential equations with random input". In: *Adv. Appl. Math. Mech.* 6.4 (2014), pp. 403–418. DOI: 10.4208/aamm.2012.m38. (Cited on page 225.)